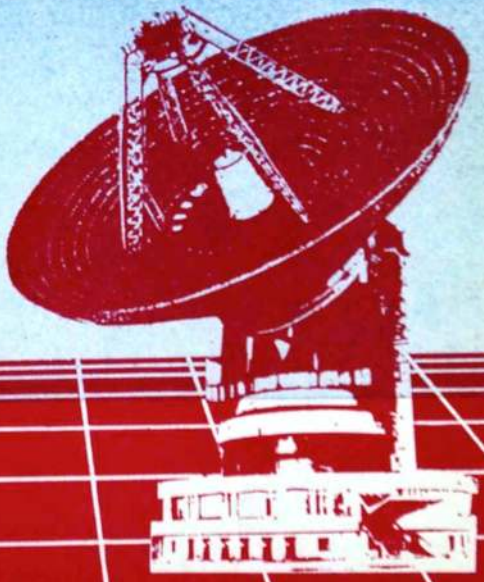
D. M. Sazonov

MICROWAVE CIRCUITS and ANTENNAS



Mir Publishers
Moscow

Д. М. Сазонов

АНТЕННЫ И УСТРОЙСТВА СВЧ

МОСКВА «ВЫСШАЯ ШКОЛА»

D. M. Sazonov

MICROWAVE CIRCUITS *and* ANTENNAS

Mir Publishers Moscow

First published 1990
Revised from the 1988 Russian edition

Translated from the Russian by
BORIS V. KUZNETSOV

На английском языке

Printed in the Union of Soviet Socialist Republics

ISBN 5-06-001149-6
ISBN 5-03-001411-X

© Издательство «Высшая школа», 1988
© English translation, Kuznetsov B. V., 1990

Contents

Preface 9

Introduction 10

Part One. Microwave Circuits 18

Chapter 1. Transmission Lines in Radio Systems and Devices 18
- 1.1. Basic Parameters of Transmission Lines 18
- 1.2. Basic Types of Transmission Lines 22
- 1.3. The Mathematical Model of a Uniform Transmission Line 36
- 1.4. The Effect of Mismatch on Transmission Line in Efficiency and Power Transfer 44
- 1.5. Impedance Transformation in Transmission Lines 48
- 1.6. Transmission-Line Charts 52
- 1.7. Narrowband Matching of Transmission Lines 54

Chapter 2. Microwave System Components 58
- 2.1. Matched Loads 58
- 2.2. Reactive Loads 60
- 2.3. Supports for Coaxial Microwave Transmission Systems 62
- 2.4. Microwave Connectors and Couplings 64
- 2.5. Transitions. Mode Transducers 68
- 2.6. Waveguide Corners, Bends and Twists 74
- 2.7. Reflecting Discontinuities in Waveguides 75
- 2.8. Mismatch in a Microwave Transmission System with Many Discontinuities 79

Chapter 3. Microwave Multiport Junctions 82
- 3.1. Basic Definitions 82
- 3.2. The Scattering Matrix 85
- 3.3. The Impedance and Admittance Matrices 91
- 3.4. Relations Between the Matrices of a Multiport Junction 98
- 3.5. Ideal and Real Matrices of a Multiport Junction 98
- 3.6. Reciprocal Multiport Junctions 99
- 3.7. Non-Dissipative Multiport Junctions 101
- 3.8. Symmetrical Multiport Junctions 110

Chapter 4. Composite Microwave Multiport Junctions 116
- 4.1. The Decomposition Principle in the Analysis of Microwave Multiport Junctions 116
- 4.2. Analysis of Cascaded Two- and One-Port Junctions by Use of Cascading Matrices 118

4.3. Decomposition of Symmetrical Four-Ports (the Even and Odd Excitation Method) 125
4.4. Directivity Kinds of Directional Couplers 129
4.5. Ring-Type Directional Couplers 135
4.6. Matched Three-Port Power Dividers 139
4.7. The Scattering Matrix of Cascaded Multiport Junctions 142
4.8. Examples of Cascaded Four- and Two-Ports 146

Chapter 5. **Microwave Filters and Matching Circuits** 150
5.1. Basic Definitions 150
5.2. Filter Prototypes with an Optimal Attenuation-vs-Frequency Response 152
5.3. Change of the Frequency Variable in the Design of Filters 155
5.4. Use of Transmission-Line Segments in Microwave Filters 159
5.5. Resonators in the Form of Discontinuities in a Transmission Line 162
5.6. Microwave Filters Using Quarter-Wave and Direct-Coupled Resonators 165
5.7. Broadband Matching of Resistive Loads by Stepped-Impedance Transformers 170
5.8. Broadband Matching of Resistive Loads by Tapered-Impedance Transformers 175
5.9. Broadband Matching of Complex Loads 177

Chapter 6. **Microwave Control and Ferrite Devices** 181
6.1. Classification of Microwave Control Devices 181
6.2. Mechanical Switching Devices. Phase Shifters. Attenuators 182
6.3. TR and ATR Tube Duplexers 185
6.4. Microwave Switching Diodes 189
6.5. Microwave Switches Using Switching Diodes 193
6.6. Discrete Switching-Diode Phase Shifters 194
6.7. Microwave Ferrite Devices 198
6.8. Nonreciprocal and Control Ferrite Devices 202
6.9. Ferrite Phase Shifters 210
6.10. Magnetically Tunable Ferrite Filters 212

Part Two. Antennas

Chapter 7. **The Electromagnetic Field and Antenna Parameters** 215
7.1. Purpose and Classification of Antennas 215
7.2. The Basic Structure of an Antenna 217
7.3. Calculation of the Electromagnetic Fields Set Up by Radiating Systems in the Far, Transition and Near Zones 219
7.4. The Complex Vector Radiation Pattern of an Antenna 226
7.5. Secondary Directional Parameters of an Antenna 234
7.6. The Transmitting Antenna as a Two-Port 238
7.7. The Bandwidth and Maximum Operating Power of an Antenna 241

Chapter 8. **The Receiving Antenna** 242
8.1. The Equivalent Circuit of a Receiving Antenna 242
8.2. Polarization Relations in Reception 246
8.3. Load Power of a Receiving Antenna. Effective Area 250
8.4. The Noise Temperature of a Receiving Antenna 251
8.5. Mutual Impedance of Spatially Separated Antennas 257
8.6. Power Transfer Between Two Antennas 259

Chapter 9. **Dipole and Slot Antennas** 263
9.1. The Electric Dipole. Current and Charge Distributions 263
9.2. The Radiation Pattern, Radiation Resistance and Directivity of a Dipole 269

9.3. Calculation of the Input Impedance of a Dipole by the Equivalent Circuit Method 274
9.4. The Symmetrical Magnetic Dipole. Slot Antennas in an Infinite Plane Screen 278
9.5. Coupled Dipoles. The EMF Method 282
9.6. The Pattern Multiplication Theorem in the Analysis of the Radiation Pattern of a System of Identical Parallel Dipoles 286

Chapter 10. Low-Directional and Frequency-Independent Antennas 290
10.1. Physical Arrangement of Dipole Antennas 290
10.2. Turnstile Antennas 296
10.3. Slot Antennas 298
10.4. Strip and Microstrip Antennas 303
10.5. Principles of Frequency-Independent Antennas 307
10.6. Frequency-Independent Spiral Antennas 310
10.7. Log-Periodic Antennas 316

Chapter 11. Linear Antennas and Arrays 319
11.1. Linear Radiating Systems. The Ideal Linear Radiator 319
11.2. Analysis of the Array Factor of an Ideal Linear Radiator 323
11.3. The Directivity of an Ideal Linear Radiator 328
11.4. The Effect of the Shape of the Amplitude Distribution on the Parameters of a Linear Antenna 332
11.5. The Effect of Phase Errors on the Parameters of a Linear Antenna 335
11.6. Analysis of the Array Factor of an Equally Spaced Linear Array. Techniques for the Suppression of the Grating Lobes 343
11.7. The Directivity of a Linear Array 349
11.8. Input Power and Gain of an Array 352
11.9. Travelling-Wave Antennas 354
11.10. Slotted Waveguide Arrays 363

Chapter 12. Radiating Apertures and Arrays 369
12.1. Basic Relationships 369
12.2. Directivity and Effective Area of a Planar In-Phase Aperture 372
12.3. The Array Factor of a Planar Aperture 374
12.4. The Equivalent Linear Radiator Method in the Analysis of a Planar Aperture 380
12.5. The Effect of Random Excitation Errors on the Parameters of a Planar Aperture 382
12.6. Beam Scanning in a Planar Aperture 384
12.7. Flat Phased Arrays. Placement of Radiators Over the Aperture and the Conditions for the Absence of Grating Lobes 387
12.8. Relationship Between the Radiation Pattern of an Array and Input Mismatch of Phased Array Elements in Scanning 392
12.9. Transition-Zone Field of an Aperture. Focused Apertures 394

Chapter 13. Synthesis of Linear Antenna Systems 401
13.1. Synthesis of Radiating Systems 401
13.2. The Radiation Pattern as an Entire Function of a Finite Degree 404
13.3. Synthesis of a Linear Radiator by the Fourier Integral Method 407
13.4. Synthesis of Difference Radiation Patterns 410
13.5. Synthesis of a Linear Radiator by the Partial Radiation Pattern Method 412
13.6. Synthesis of Linear Arrays 415
13.7. Supergain Antennas 420

Chapter 14. Aperture-Type Antennas 422
14.1. General Properties 422
14.2. Horn Antennas 423
14.3. Lens Antennas 427
14.4. Parabolic Reflector Antennas 435
14.5. Feeds for Reflector Antennas 444
14.6. Special Types of Reflector Antennas 450

Chapter 15. Arrays 459
15.1. Scanning Methods. Objectives Served by Arrays 459
15.2. Phased Arrays 461
15.3. Phasing Control of Electronically Scanned Arrays 467
15.4. Multibeam Arrays 471
15.5. Frequency-Scanned Arrays 477

Conclusion 481

Appendix. Electrodynamic Foundations of Antenna and Microwave Component Theory 484

Bibliography 497

Index 498

Preface

The objective of this text is to give the reader an insight into the theoretical aspects, analysis, synthesis and design of antennas and their r.f. transmission (feeder) systems, including preparatory steps for the use of existing and contemplated CAD systems. It is presumed that the reader has a working knowledge of the most generally used divisions of higher mathematics, basic circuit theory, and Maxwell's equations. From a course on electrodynamics, the reader is supposed to know the characteristics of guided and unguided electromagnetic waves, the fundamental principles of reciprocity, duality, equivalence, and the Poynting theorem. As a reminder, these topics are recapitulated in the appendix to the text.

Basically, this text follows in the steps of Antennas published by G.T. Markov and D.M. Sazonov in 1975, and Microwave Circuits by D.M. Sazonov, A.N. Gridin and B.A. Mishustin which appeared in an English translation in 1981. In contrast to the above two books, a discourse on theory is combined with practical designs, and there is no division of the material into two separate parts. It is to be stressed that because of the limited space available, there are no separate sections on the measurement of characteristics of microwave circuits and antennas, including antennas for short, medium and long wavelengths. The relevant theoretical aspects, however, will be found in other chapters to the extent envisaged by the curriculum. For a more detailed exposition of the subject-matter the reader may turn to [5] and [6], and also Chapters 15 and 16 in [1].

The Author

Introduction

Any radio (communication, broadcasting, radar, navigation and other) systems will always include each as its indispensable parts an antenna and an r.f. transmission (or feeder) system.

For better insight into the roles played by antennas and r.f. transmission (feeder) systems, let us refer to the functional diagram of a radio communication system shown in Fig. 0.1. The transmitting antenna transforms the guided electromagnetic waves coming from

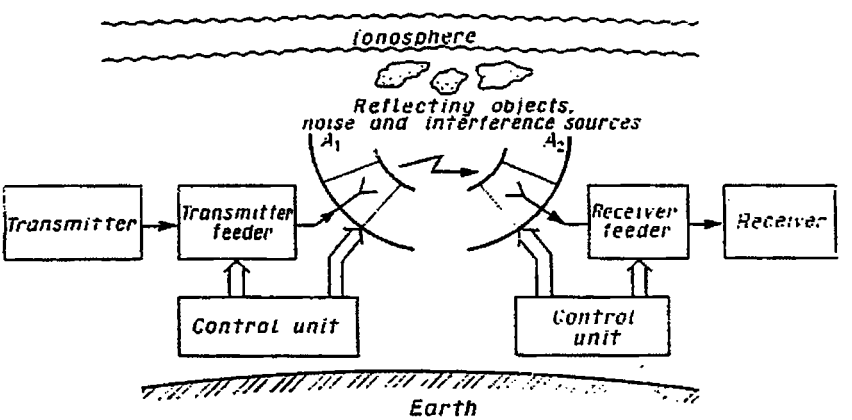


Fig. 0.1. General arrangement of a radio system

the transmitter over the r.f. transmission system into unguided electromagnetic waves which are then radiated into free space. On its part, the receiving antenna transforms the incoming unguided (free-space) electromagnetic waves into guided waves which are then propagated in the receiving r.f. transmission system. This conversion of unguided into guided electromagnetic waves is inevitably accompanied by the reverse emission of some radiation. The reverse radiation field due to the receiving antenna is superimposed on the primary field of the transmitting antenna in such a way that the overall power

of the two fields carried to infinity is reduced by the power that the receiving antenna extracts from the incident field. There is no fundamental difference in the construction of a transmitting and a receiving antenna. According to the reciprocity principle known from the theory of electromagnetism, any antenna can be used for both transmission and reception. This explains why in many radio systems, radio waves will usually be transmitted and received by a common antenna.

State-of-the-art antennas are designed to meet a number of requirements. Of them, two are decisive. One is directive properties or the distribution of electromagnetic power in space (or the response to the incoming electromagnetic field during reception) in a well-defined pattern. Sometimes, it is desired that an antenna should radiate or receive equally well in all directions. In other cases, it is required that an antenna should radiate or receive preferably within some narrow sector called the beam. For an antenna to produce a narrow beam, its dimensions should be many times the operating wavelength of the associated radio system. The other key requirement is that the emission and reception of radio waves should be accompanied by the least attainable loss of electromagnetic power through dissipation of heat in the conductors and dielectrics of the antenna; that is, an antenna should have a high efficiency. The task of achieving a high efficiency is especially formidable in cases where an antenna is very small in comparison with the operating wavelength.

How a practical antenna will operate under actual service conditions and how one should go about choosing the operating frequencies (wavelengths) and antenna parameters for a given radio system strongly depends on the properties of the medium in which the waves are to be propagated. Traditionally, these matters are dealt with in a separate course on the electrodynamics and propagation of radio waves. In this text it is assumed that antennas operate under the idealized conditions of free space. The properties and effects of the surrounding medium can be accounted for by recourse to the methods offered by the theory of wave propagation.

To a very important extent the performance of a radio system depends on the r.f. feeders that connect its antennas to its transmitter and receiver (see Fig. 0.1). Each r.f. transmission system channels electromagnetic waves as appropriate, maintains the right conditions for the operation of the output and input circuits in the transmitter and receiver, and pre-selects the incoming signals on a frequency basis. It may contain switching circuits and rotary (or rotating) joints, electronic-beam switching facilities, beam-polarization devices, and whatever instrumentation may be necessary.

Apart from antennas, r.f. transmission systems are widely used in measuring equipment, notably to determine the parameters of

various media, in particle accelerators, in the heating and drying of materials and products, medicine and elsewhere. Therefore, we will define an *r.f. transmission system* in a wider sense *as an assemblage of r.f. or microwave devices coupled in a particular manner to achieve a particular task.*

Most commonly, the key elements of an r.f. transmission system are transmission-line (or waveguide) sections, adapters, connectors, matching elements, directional couplers, absorbers, filters, phase shifters, ferrite circulators and isolators, and switches.

Predominantly, antennas and the associated r.f. transmission systems fall in the category of linear and passive devices. Physically, they are assemblages of components and elements made of conductors, dielectrics, and magnetodielectrics.

Recent years have seen a growing emphasis being placed on the devices that are intended to control the operation of antennas and r.f. (especially microwave) transmission systems (see Fig. 0.1), especially in the case of antennas using a beam rapidly moved in space. Usually termed phased arrays, such antennas are commonly built as a system made up of a great number of individual radiators, with the phase of the r.f. excitation for each being independently controlled by fast-acting semiconductor or ferrite control devices in response to commands from a computer. Matters related to the design of control devices for antennas and r.f. or microwave transmission system fall in the field covered by courses in telecommunication systems and radio automation and will not therefore be taken up here. When presenting the theory of antennas, however, mention will be made about the ways and means of controlling the spatial position of the antenna beam. Also, in discussing the elements that make up microwave transmission systems, devices will be touched upon for controlling the amplitude and phase of excitation for phased array radiators.

The emission and reception of radio waves by antennas and the propagation of electromagnetic waves in r.f. transmission systems and their components involve highly complicated *wave processes.* An adequate mathematical description of such processes is offered by the general theory of the electromagnetic field (electrodynamics) based on Maxwell's equations supplemented with material equations for media and boundary conditions.

Although they appear outwardly simple and physically straightforward, the direct use of Maxwell's equations in the design of specific antennas and transmission systems mostly fails to yield the desired results because of appreciable mathematical handicaps. As has been found, exact and complete solutions of electrodynamic problems even for the simplest antennas (such as isolated dipoles and slot radiators) and standard elements of microwave transmission systems (such as abrupt changes in line's dimensions, irises in waveguides, or arms) yield prohibitively involved representations of electric

and magnetic field intensities as vector functions of three spatial coordinates. Fortunately, it is not at all necessary to reconstruct a complete picture of the electromagnetic field at any point in space when designing most antennas and r.f. or microwave transmission systems. What is important is to be able to determine and assure within the specified limits the desired performance of the antenna (its radiation pattern, input impedances, and the like) and an adequate response of the transmission systems to a particular excitation.

In calculating the electrical characteristics of antennas or transmission systems, it has been proved successful to use, in addition to a rigorous solution of boundary problems, far simpler engineering approaches. Those to be mentioned in the first place are the general theory of microwave circuits, based on the matrix techniques of linear algebra, the theory of the function of a complex variable, and some other divisions of mathematics. One of the pivotal ideas of the general theory of microwave circuits is that any transmission line or waveguide may be replaced by equivalent distributed-parameter (or distributed-constant) long lines. The remaining elements of a microwave transmission system or an antenna (such as waveguide junctions, loads, irises, radiating slots, dipole arrays, and so on) are replaced for purposes of analysis and synthesis by equivalent multiport junctions. These multiport junctions are then described either by parameter matrices (such as scattering, impedance or transfer matrices) or by equivalent circuits of an adequate level of complexity. The parameters of the equivalent multiports or of their equivalent circuits are deduced subject to the condition that the electrical characteristics found for a particular element from its equivalent circuit should correspond in a specified frequency band to the exact values obtained by a rigorous solution of the respective boundary problem of electrodynamics or by precision measurements.

If an antenna or transmission-system designer has at his disposal a sufficiently ample collection of equivalent circuits for *standard*, or *base*, *elements*, he may then proceed with his analysis or synthesis task by resort to algorithms that combine the base elements into a complete system of the desired configuration. With such an approach it is not any longer necessary to know in detail how electromagnetic fields behave in the various devices, and the combining algorithm itself is implemented by a computer which executes suitable programs developed and debugged in advance. The package of such programs along with a bank of base element characteristics constitutes the kernel of a *CAD system* for a particular class of antennas and transmission systems.

In addition to the general theory of microwave circuits, designers of state-of-the-art antennas and transmission systems widely use other approximate theories, such as geometrical optics, physical

(wave) optics, the scalar theory of diffraction, and the geometrical theory of diffraction.

A Historical Overview. The science and art of antennas and r.f. transmission (or feeder) systems, as we know them today, go back to the 19th century. As tradition has it, the early serious scientific ideas related to the electromagnetic field are associated with the experiments carried out by M. Faraday (1791-1867). A rigorous mathematical foundation for electromagnetism was laid in 1864 by J. Maxwell (1831-1879) as a set of universal equations. Following that, the most significant theoretical and experimental studies into the structure of the fields set up by the elementary dipole and other simple electromagnetic wave radiators were performed by H. Hertz (1857-1894), but he failed to see anything of practical value in his findings. It was not until 1895 that A.S. Popov (1859-1906) built — for the first time ever — a transmitting antenna (as square metal sheets secured at the ends of a Hertzian dipole) and a receiving antenna (as a vertical conductor and a grounding arrangement). The theoretical treatment of a dipole antenna as an array of dipoles was proposed by M. Abraham who formulated the concept of the radiation resistance of an antenna in 1900.

That electromagnetic waves could be transmitted within metal pipes was first suggested theoretically by J.J. Thomson in 1893 and Lord Rayleigh in 1897. It was not until 35 years later that their ideas were implemented in practice.

Microwave antenna and circuit technology took a firm root in 1940-50s. Arbitrarily, the following phases might be defined in the development of antennas and r.f. transmission systems.

1. The period of wire antennas for long and medium waves (1900-1925). The antennas were small in comparison with the wavelengths used, and the principal bottleneck was to assure an acceptable radiation efficiency. The best antennas of the time used masts up to 150 m tall and ramified grounding systems. A very fruitful idea was advanced by Alexanderson who proposed the use of a vertical antenna with several in-phase tuned downleads and a well-developed horizontal part. In 1917-1918, M.V. Shuleikin came up with a method for calculating the capacitance and inductance of complex long-wave radio networks.

2. The period of short-wave antennas (1920-1935). As radio communication moved into the short-wave band, antennas came to be used which were not only comparable in size but even could substantially exceed in that respect the operating wavelength. This offered an opportunity to build directional antennas. Among the many accomplishments of this period, mention should be made about in-phase multi-dipole antennas (the predecessors of the present-day phased arrays). In the USSR, they were developed under the guidan-

ce of M.A. Bonch-Bruevich and V.V. Tatarinov (for the Moscow-Tashkent radio link which went into operation in 1926).

3. The next period is one of advances in the use of ultrashort waves. In the Soviet Union, a major effort was launched by B.A. Vvedensky and A.I. Danilevsky in 1921. However, ultrashort-wave antennas and their transmission systems came to be regularly used between 1930 and 1940. It was then that reliable sources of CW oscillations in the decimetric (UHF) and centimetric (SHF) bands were developed, thus promoting work on practical waveguides and stimulated the search for better designs of waveguide components. Advances in television and ultrashort-wave broadcasting underscored the need for broadband antennas. Early designs were developed for horn antennas and quasi-optical antennas (lens and reflector types).

4. The art and science of antennas and microwave circuits between 1940 and 1960 were greatly affected by the advent of UHF and SHF radar systems. Those years saw a good deal of effort put into the theory and engineering of the most commonly used highly directional antennas — the reflector, horn, and lens types. Much ground was gained by a wide range of slot antennas (both resonant and arrays of rectangular waveguides) and extrabroadband types (log-periodic and log-spiral antennas). Between 1950 and 1960, the final shape was given to the theory of passive rectangular and coaxial waveguide components. Also, nonreciprocal ferrite devices and controlled ferrite phase-shifters came on the stage, and probing steps were made in the art and science of strip transmission lines and dielectric waveguides.

5. The current period in the development of antennas and microwave circuits is closely associated with several milestones in history: space exploration by manned missions unthinkable without an adequate telecommunication support, explosive advances in computers on the basis of integrated circuitry, the practical use of millimetric and optical wavelengths, and the advent of strip, microstrip and fibre-optic transmission lines — all this has served to reduce the size and to improve the performance of microwave transmission systems and the associated antennas.

As regards antennas, the most impressive feats in this period have been the advent of practical electronically scanned phased arrays (in which the beam scans the assigned sector in a matter of a few microseconds or even several fractions of a microsecond), space-deployable pencil-beam reflector-type antennas and arrays for radio communication and broadcasting and Earth resource surveying, and gigantic allround-looking antennas with reflectors 60 to 100 m in diameter for radio astronomy and deep-space radio communication.

It has taken the strenuous effort of thousands of scientists and engineers in many countries to set up the theory and art of present-day microwave circuitry and antennas. Quite a number of valuable

contributions have come from the Soviet Union. As early as 1927, A.A. Pistolkors formulated a basic theory of single- and multi-wire transmission lines. In 1930-1940, his theory was carried an important step forward by V.V. Tatarinov who proposed the now most commonly used method for measuring complex input impedances in transmission lines from the longitudinal distribution of voltage or current and devised some of the now classic circuits to match loads to transmission lines.

In 1939, A.A. Pistolkors and M.S. Neyman formulated theoretical foundations for and developed the early designs of directional couplers — the key 'building blocks' of present-day ramified microwave transmission systems. In 1940, A.R. Volpert came up with the idea of using directional couplers as reflectometers — microwave systems arranged to measure the incident and reflected voltages and to indicate their ratio, or the reflection factor. A year before, he proposed a circle diagram for transmission lines — it materially simplified design work on transmission lines and matching devices.

In 1940-1950, G.V. Kisunko and G.T. Markov were the first to work out important theoretical aspects of waveguide excitation. In the 1960s, A. L. Mikaelyan formulated the theory of nonreciprocal ferrite devices. In the 1970s, B.V. Sestroretsky proposed a theory of switching and phasing devices built around semiconductor diodes. In 1960-70s, A.L. Feldshtein proposed a theory for the synthesis of filters and directional four-ports.

In the field of antennas, too, Soviet contributions have been no less important. Multi-element antennas are still designed by the EMF method based on works by D.A. Rozhansky, I.G. Kliatskin, A.A. Pistolkors and V.V. Tatarinov, which go back to the years 1922-1928. In 1928, A.A. Pistolkors formulated the concept of the directivity of an antenna, now generally accepted by all those working in this field. In 1935, M.S. Neyman was the first to apply the principle of reciprocity to receiving antennas, thus laying the foundation for the theory of receiving antennas in general. M.S. Neyman was also the first to propose the idea of a slot antenna in 1938. The basis for a theory of such antennas was laid by A.A. Pistolkors who formulated, in 1944, the principle of duality which establishes correspondence between the characteristics of wire and slot antennas. The theory of slot antennas was finally formulated by Ya.N. Feld in 1945-48. In 1958, L.A. Vainshtein obtained an exact electrodynamic solution for the problem of radiation from the open end of a waveguide — this had an important bearing on the theory of aperture-type (notably, horn) antennas.

Soviet contributions have been especially valuable as regards the theory of synthesis of antennas which have a specified radiation pattern. The early fundamental results in this field were obtained by A.I. Uzkov as early as 1945. This theory was carried forward by

L.D. Bakhrakh, Ya.N. Feld, Ye.G. Zelkin, V.P. Yakovlev, and other scientists.

Soviet scientists have also been trail-blazers in the use of computers for electrodynamic calculations of complex radiating systems and antennas. Notably, Ye.N. Vasilyev used a method of integral equations as early as 1960 to calculate, with the aid of a computer, the radiation characteristics of antennas placed near metal bodies of revolution. In 1970-1980, V.V. Nikolsky and his co-workers proposed several ingenious methods for computer-aided electrodynamic analysis and synthesis of waveguide and microstrip devices and transmission lines.

Among the results obtained by Soviet scientists in recent years, special mention should be made about the use of computers for the recovery of the radiation characteristics of large-sized antennas from the measured amplitudes and phases of the near electromagnetic field. Credit for pioneer work here goes to P.M. Geruni and L.D. Bakhrakh.

In summing up this short historical overview of antennas and microwave circuits, we may note with satisfaction that from a simple means solely intended to extend the range of radio communication, as it was used in Popov's early devices, the antenna has matured to become the heart of any radio system. The range and accuracy of present-day radars, the sensitivity and resolution of radio telescopes, the range of space communication and many other parameters of various radio systems are all determined by the parameters of antennas achievable at the present state of the art, notably the beam width, or directional properties. Combined with multi-channel transmission (or feeder) systems, sophisticated antennas have functionally become an electronic analog of the eye capable of 'radio vision'.

Part One

Microwave Circuits

Chapter One

Transmission Lines in Radio Systems and Devices

1.1. Basic Parameters of Transmission Lines

A transmission line* refers to a material structure forming a continuous path from one place to another and used to direct electromagnetic energy from its source to loads. Transmission lines serve to convey power from generators to using equipment, to form resonant systems such as resonant cavities and distributed-parameter tuned circuits, and to transform load impedances. Transmission-line sections are used to combine the individual microwave components into an integral r.f. or microwave transmission system.

A transmission line is called *uniform* if its cross-sectional area and the electromagnetic properties of the filling medium (or media) do not change with distance along its axis. Also, a transmission line is called *homogeneous* if its cross-sectional area is filled by a homogeneous medium. A further subdivision is into *open lines* and *waveguides*. An open line (such as multi-wire and strip lines, surface-wave lines, and some others) has no outer conducting surface that would completely enclose the conducting region. In contrast, a waveguide will have one or several conducting surfaces completely enclosing the region in which electromagnetic waves are propagated.

When choosing a particular type of transmission line for a specific application, the following electrical characteristics and parameters are decisive.

Wave type or mode, field pattern, and cutoff frequency. To find these characteristics, one proceeds from the specified cross-sectional

---

* In Soviet usage, the term 'transmission line' covers both transmission lines proper and waveguides. This seems to be in contrast to the usage adopted by many authors outside the USSR. In this text, Soviet usage will mostly be adhered to.— *Translator's note.*

area and solves the boundary problems of electrodynamics. Most frequently, electromagnetic energy is conveyed on a transmission line by the *dominant* (or *principal*) *type of wave*, or *mode*, that is, one having the lowest *critical* or *cutoff frequency*, ω_c on a given line. In some applications, however, preference may be given to *higher modes* for which the cutoff frequency exceeds that of the dominant mode.

The following types of waves, or modes, may be named.

(a) *Transverse electromagnetic waves*, or the TEM *mode*. In a homogeneous isotropic medium, its electric and magnetic field vectors are both everywhere perpendicular to the direction of propagation. In other words, the longitudinal components of both field vectors are zero; the cutoff frequency for the TEM mode is likewise zero. It exists solely in lines having at least two isolated conductors.

(b) *Electrical* (*E*) *waves*, or the TM (*transverse magnetic*) *mode*. In this mode the longitudinal component of the magnetic field is zero and the longitudinal component of the electric field is non-zero.

(c) *Magnetic* (H) *waves*, or the TE (*transverse electric*) *mode*. In this mode the electric field vector is everywhere perpendicular to the direction of propagation, and the longitudinal component of the electric field vector is zero.

(d) *Hybrid electromagnetic* (HE or EH) *waves*, or the HE *mode*. This mode has components of both electric and magnetic field vectors in the direction of propagation.

Both TE and TM modes are typical of waveguides filled with a homogeneous, isotropic medium. Their cutoff frequencies are other than zero and depend on the shape and dimensions of the cross section and the parameters of the filling dielectric. Exact equations have been derived for the cutoff frequency or cutoff wavelength of waveguides with a regularly shaped cross-section (a rectangle, a circle, and so on).

Hybrid electromagnetic waves exist, as a rule, in transmission lines using an inhomogeneous dielectric medium. Their cutoff frequencies depend on the shape and size of the cross-section and the parameters of the filling medium in a complex manner. Elaborate algorithms have to be used for design in such cases, and so the characteristics of hybrid electromagnetic waves are usually found on a computer. For some hybrid modes, the cutoff frequency may asymptotically tend to zero.

Dispersion. If a transmission line or a transmission medium is such that different frequencies travel at different velocities, the line or the medium is said to be *dispersive*. This dependence of velocity on frequency for a particular line may be stated as an equation or a plot.

Transmission lines operating in the TEM mode are free from dispersion, and the phase velocity in them at any frequency is equal to the velocity of propagation of a plane electromagnetic wave in the

medium filling the transmission line:

$$v_p = c\,(\varepsilon_r\mu_r)^{-1/2}$$

where c = velocity of light in vacuum
ε_r = relative permittivity of the medium
μ_r = relative permeability of the medium

The phase velocities of TE and TM waves excited in a hollow waveguide are frequency-dependent and exceed the velocity of light in the medium filling the waveguide. The respective relation may be expressed by a general equation

$$v_p = c\,(\varepsilon_r\mu_r)^{-1/2}\,[1 - (\lambda_0/\lambda_c)^2]^{-1/2}$$

where λ_c = cutoff wavelength of the mode, taking into account the properties of the medium filling waveguide
λ_0 = wavelength of the mode in vacuum

Accordingly, for TE and TM modes the wavelength in a hollow waveguide (the guide wavelength) will be

$$\lambda_g = \lambda_0\,(\varepsilon_r\mu_r)^{-1/2}\,[1 - (\lambda_0/\lambda_c)^2]^{-1/2}$$

For hybrid modes, the phase velocity depends on frequency in a more complex way, and for them the dispersion characteristic has to be found on a computer, using suitable programs.

Dispersion has a marked bearing on the frequency response of devices built of transmission-line (waveguide) sections, and should be taken into account when designing such devices.

Attenuation constant. An electromagnetic wave propagated in any real transmission line suffers attenuation: the magnitude of its power decreases as

$$P\,(\xi) = P\,(0)\,\exp^{-2\alpha\xi}$$

where $P\,(0)$ = power carried by the travelling wave through the reference (initial) section of the transmission line
α = attenuation constant
ξ = distance travelled by the wave in the direction of propagation

Attenuation takes place because some of the propagated power is inevitably dissipated as heat in the conductors and dielectrics, and also because some power is radiated.

The attenuation constant may be expressed either in nepers per metre of line length

$$\alpha = 0.5\,\ln\,[P\,(0)/P\,(1)]$$

or in decibels per metre of line length:

$$\alpha = 10\,\log\,[P\,(0)/P\,(1)]$$

The two units are connected by the following obvious relations:

$$1\ \text{dB/m} = \text{approx. } 0.115\ \text{neper/m}$$

and

$$1 \text{ neper/m} = \text{approx. } 8.68 \text{ dB/m}$$

The calculation of the attenuation constant for transmission lines is discussed in detail in courses on electrodynamics. Here it will suffice to recall that the most commonly used equation for the attenuation constant (in nepers per metre of line length) has the form

$$\alpha = P_1/2P(0)$$

where P_1 is the power lost per metre of line length.

In determining power loss per unit length, it is usually assumed that the field pattern and the current distribution over the surfaces of the line's conductors are the same as they would be if there were no losses. As the theory of skin effect tells us, the skin depth δ (cm) and the resistance presented by a unit square of the conductor surface R_s (ohms) are given by

$$\delta = 3.8 \times 10^{-5} (\lambda_0/\mu_m)^{1/2}$$

and (1.1)

$$R_s = 0.045 A (\mu_m/\lambda_0)^{1/2}$$

where μ_m = relative permeability of the metal
λ_0 = generator wavelength, cm
$A = (\sigma_{Cu}/\sigma_m)^{1/2}$ = factor taking care of the difference in conductance between the conductor metal and that of copper (it is 1 for copper, 0.98 for silver, 1.35 for aluminium, 2.1 for brass, and 1.2 for gold).

The power lost per unit length of a transmission line can be found by taking the integral of the elementary losses, $0.5 \mid J_m \mid^2 R_s$, over the perimeter of the cross section of the line conductors. Here, J_m is the surface current density.

In calculating power dissipation in the line dielectric, it is assumed that the pattern of the electric field E_m in the line is undisturbed, and the elementary dissipation, $0.5 \mid E_m \mid^2 \omega \varepsilon_r \tan \delta_d$, is integrated over the cross-sectional area of the dielectric. Here, ε_r is the relative permittivity of the dielectric filling the line's segment in question, and $\tan \delta_d$ is the loss tangent to the dielectric.

A rise in frequency nearly always leads to an increase in attenuation. There are a few exceptions, such as for the TE_{01} mode in a circular waveguide. The attenuation due to losses in the dielectric increases in direct proportion to frequency ω (with the loss tangent $\tan \delta_d$ held constant), while the attenuation due to losses in the line conductors rises at a lower rate — in direct proportion to the square root of frequency. In hollow waveguides, the attenuation is further dependent on the ratio between the operating and cutoff frequencies: as the

operating frequency approaches the cutoff value, the attenuation builds up at a progressively higher rate so that it tends to infinity at the cutoff frequency.

Maximum transmitted power. Under travelling-wave conditions, the maximum transmitted power is limited by the electric breakdown or the excessive heating of the conductors and insulators of the line. In pulsed working at high values of pulse period to pulse duration ratio (a low pulse duty factor), the electric breakdown is more dangerous, whereas in CW work involving large blocks of power the line may experience a thermal breakdown. The maximum allowable power for a line is usually set at 25 to 30 % of the critical power that would cause a breakdown or heat build-up under conditions of a purely travelling wave. It is customary to allow a three-fold margin of safety as a precaution against any likely impairment in electrical strength due to various discontinuities and mismatch in the r.f. or microwave transmission system.

As the operating wavelength is reduced, the cross-sectional dimensions of the line have also to be reduced so as to prevent the excitation and propagation of undesired higher modes. A decrease in the cross-sectional dimensions entails an increase in the field concentration in the transmission line and a proportionate decrease in the maximum transmitted power (power rating). As already noted, at shorter wavelengths the attenuation goes up; coupled with an impaired heat withdrawal due to the reduced cross-sectional area, this likewise sets a lower limit for the maximum transmitted power because of the likely heat build-up.

1.2. Basic Types of Transmission Lines

The types of transmission lines used in communication systems are many and diverse. The choice of a particular type depends on the purpose to be served by a given system, its operating conditions and, decisively, by the wavelengths to be used and the power to be transmitted. The most commonly used types of transmission lines are shown in Fig. 1.1*. Let us take a closer look at some of them.

Wire transmission lines. Open lines made up each of two identical parallel wires (Fig. 1.2*a*) and operated in the TEM mode are used in the hectometric (ME) and metric (HF) bands and serve to connect antennas to receivers and transmitters. On shorter wavelengths, the use of two-wire lines is limited by an appreciable radiation which interferes with the operation of neighbouring equipments and increases signal attenuation.

The characteristic impedance of an overhead two-wire line (in ohms) depends on the conductor diameter d and the conductor spacing

* See the footnote on page 18.— *Translator's note.*

D as stated in the following equation:

$$Z_c = 276 \log \{D/d + [1 + (D/d)^2]^{1/2}\}$$

With the values of D/d chosen to be convenient from structural considerations, the characteristic impedance of a two-wire line is 200 to 600 ohms.

Overhead two-wire lines are built from bare copper or composite metal wires suspended from supports by specially designed insulators at no less than 3 m above the ground.

Four-wire lines made from pairwise connected conductors (Fig. 1.2*b*) are used in the same applications as their two-wire counter-

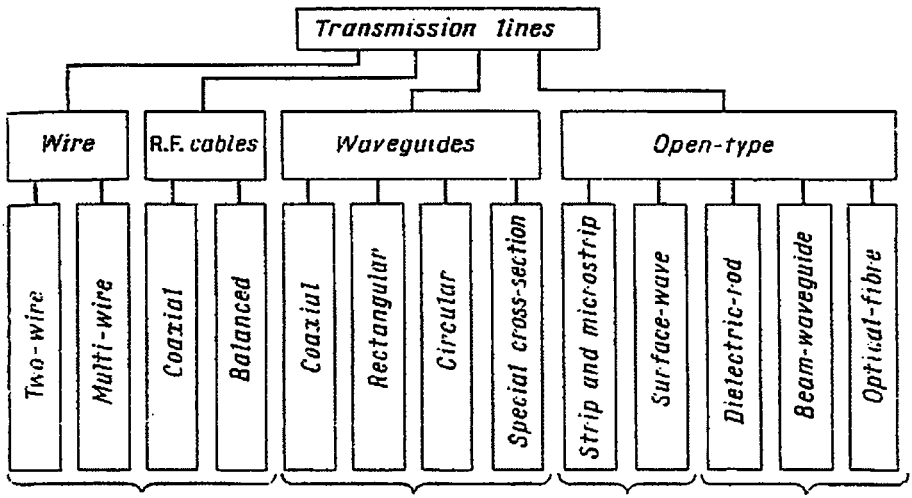


Fig. 1.1. Types of transmission lines for various wavelengths

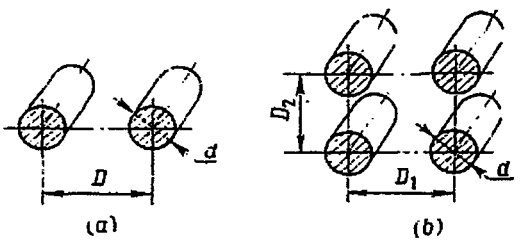


Fig. 1.2. Wire transmission lines operating in the TEM mode

parts, but emit a smaller amount of spurious radiation, have a lower characteristic impedance, and show a better electric strength.

A four-wire line can be formed in any one of two ways: by arranging the in-phase pairs of conductors either in a vertical parallel pattern

or in a cross-wise opposite (transposed) pattern. With in-phase (or even) excitation of the wires lying on one side of the vertical plane of symmetry, the characteristic impedance is given (in ohms) by

$$Z_{c1} = 138 \log \{(2D_1/d)\,[1 + (D_1/D_2)^2]^{1/2}\}$$

If even (in-phase) excitation is applied to the wires lying crosswise opposite to one another, the characteristic impedance is reduced to a value given by

$$Z_{c2} = 138 \log \frac{2D_1}{d\,[1+(D_1/D_2)^2]^{1/2}}$$

The typical dimensions for a four-wire transposed line used as a feeder for transmitting antennas are as follows: $D_1 = D_2 =$ approx. 40 cm, $d =$ approx. 5 mm, and the characteristic impedance is 300 ohms. More compact lines are used to connect antennas to receivers, with $D_1 = D_2 =$ approx. 3 cm, $d =$ approx. 1.5 mm, and a characteristic impedance of 200 ohms.

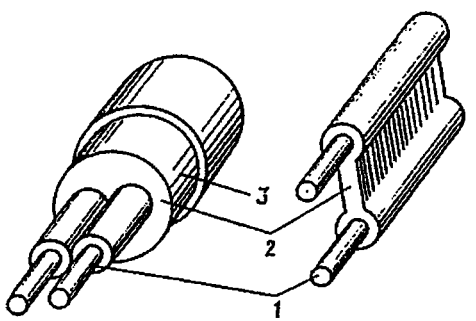


Fig. 1.3. Unshielded two-wire cables: *1*, line conductors; *2*, dielectric; *3*, dielectric sheath

The attenuation (in nepers per metre run) due to losses in the line conductors is found by the general equation

$$\alpha_m = R_1/2Z_c$$

where R_1 is the loop resistance per unit length (per metre run), equal to the sum of the resistance per metre run of the individual conductors.

Two-wire lines have a relatively low attenuation. For a line using copper conductors such that $D = 70$ mm and $d = 1$ mm, the characteristic impedance is $Z_c = 600$ ohms, and the design attenuation at a wavelength of 30 m is 0.005 dB/m.

Sometimes, two-wire lines are also fabricated in flexible cables, as shown in Fig. 1.3.

Coaxial transmission lines. These lines are used at wavelengths from 3-5 cm to 10 m. They may be coaxial waveguides made of rigid metal pipes so that one is supported in the other by dielectric discs or metal insulators, or flexible coaxial coables.

The characteristic impedance of a coaxial waveguide operating in the TEM mode (in ohms) is given by

$$Z_c = 138\,(\mu_r/\varepsilon_r)^{1/2} \log (D/d)$$

In the general case, the signal attenuation (in dB/m) in a coaxial line is caused by losses in the conductor and dielectric, so the total

attenuation is the sum of two components

$$\alpha = \alpha_m + \alpha_d$$

Each component can be described by the following equations:

$$\alpha_m = \frac{2.3R_s(1+D/d)}{D\ln(D/d)}(\varepsilon_r/\mu_r)^{1/2}$$

and (1.2)

$$\alpha_d = \frac{2730\varepsilon_r^{1/2}\tan\delta_d}{\lambda_0}$$

where R_s is the resistance presented by a unit square of the conductor surface (in ohms), as given by Eq. (1.1). The diameter D and the wavelength λ_0 are expressed in centimetres.

Note that the value of α_d is independent of the cross-sectional geometry of the line and is solely decided by the parameters of the dielectric and the operating wavelength (which holds for any lines operating in the TEM mode). In contrast, the value of α_m depends on the conductor geometry, so it is worth while finding out which value of D/d reduces α_m to a minimum. A simple analysis would show that α_m is a minimum at $D/d = 3.6$, which corresponds to a characteristic impedance of 77 ohms for an air-filled line. However, the ratio $\alpha_m/\alpha_{m\,min}$ varies with the value of D/d in a rather gradual manner (Fig. 1.4).

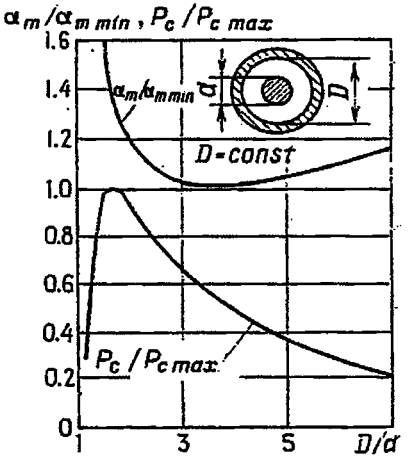


Fig. 1.4. Characteristics of a TEM-mode coaxial waveguide

The critical power for a coaxial waveguide (that is, one at which an electric break-down is about to happen) operating in a purely travelling wave mode is given by

$$P_c = (E_c^2/240\pi)(\pi D^2/4)\frac{2\ln(D/d)}{(D/d)^2} \quad (1.3)$$

where

$$E_c^2/240\,\pi = 1.2\ \text{MW/cm}^2$$

is the magnitude of the Poynting vector at the onset of a breakdown of air by a uniform plane wave, and $\pi D^2/4$ is the cross-sectional area of the coaxial waveguide. The shape factor

$$\frac{2\ln(D/d)}{(D/d)^2}$$

in Eq. (1.3) accounts for the nonuniform distribution of the electric field over the cross section of the waveguide. It is a maximum at $D/d = 1.65$, being 0.368 in value, when the characteristic impedance is 30 ohms. As the ratio D/d departs from its optimal value, the relative electric strength $P_c/P_{c\,max}$ of a coaxial waveguide decreases at a progressively higher rate (see Fig. 1.4).

As follows from Eqs. (1.2) and (1.3) which define the attenuation and the breakdown (critical) power under conditions of an electric break-down, the value of α_m could be decreased and the electric strength could be enhanced by increasing the waveguide diameter D (rather, its cross-sectional area). This might, however, give rise to the TE_{11} mode for which the cutoff wavelength is

$$\lambda_c \approx \pi (D + d)/2.02$$

If instead of $D = \text{const}$, we require that $\lambda \geqslant \lambda_c$ (which for the TE_{11} mode means a constant sum of the outside and inside diameters of the line conductors), the relative attenuation and the relative electric strength will vary with D/d in the same manner as shown in the plot of Fig. 1.4, but the extrema will take up different positions. The signal attenuation is a minimum at $D/d \approx 4.67$, and the electric strength is a maximum at $D/d \approx 2.1$. The most commonly used standard characteristic impedances of 50 ohms and 75 ohms have been chosen in a region where the parameters of coaxial waveguides are close to optimal.

As has been noted, there may also be *flexible coaxial cables.* A flexible coaxial cable consists of a single- or a multi-wire inner (or central) conductor and an outer conductor in the form of metal braiding, the two being separated by an elastic dielectric (such as polyethylene or PTFE). For protection against external influences, the cable is additionally covered by a dielectric sheath or jacket. The parameters of coaxial cables can be looked up in reference sources.

Strip and microstrip transmission lines. Widely used in the decimetric (UHF) and centimetric (SHF) bands, these lines mainly serve to interconnect various microwave devices and components into a ramified microwave transmission (or feeder) system.

A common type of strip line consists of a thin central conductor of rectangular cross-section situated midway between two parallel metal plates ('ground planes'), with the long dimension of the central strip conductor running parallel to the outer plates. Because the central strip is placed between two parallel plates arranged symmetrically about the strip, this form of transmission line is variously called *shielded*, *balanced*, or *symmetrical.* There is another form, known as an *open-strip line* which has only one ground plane. Sometimes it is referred to as an *unshielded*, *unbalanced* or *unsymmetrical strip*. The more common usage is to call the former type simply as strip lines, and the latter, as microstrip lines. Several forms of

strip and microstrip lines are shown in Fig. 1.5, along with the associated electric field patterns over the cross-section.

Strip and microstrip lines may come in any one of three designs: rigid air-filled lines, lines fabricated on foil-cladded dielectric substrate sheets (glass-reinforced PTFE, polymers, and the like), and lines fabricated on dielectric substrate sheets of ceramic or crystalline materials of high permittivity (pyrosil, quartz, sapphyre, silicon, gallium arsenide, and the like).

Rigid air-filled strip lines are mostly of shielded (or balanced) construction (Fig. 1.5*b*) and are used in applications involving high

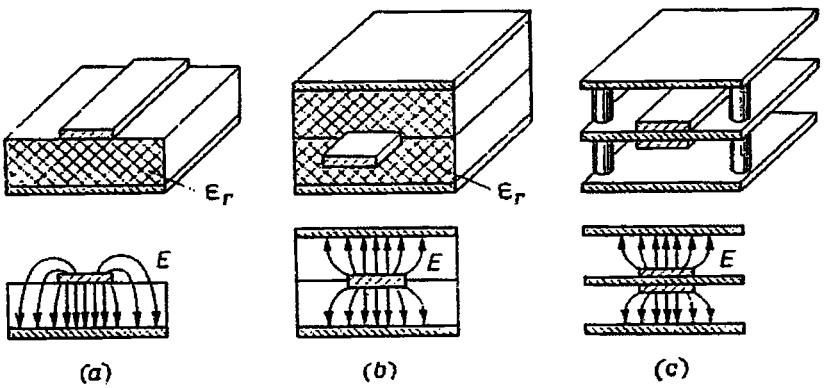


Fig. 1.5. Microwave strip lines: *a*, microstrip; *b*, triplate line; *c*, high-*Q* strip line

levels of microwave power. For better rigidity, the central conductor is fabricated from a heavy-gauge strip, and is held in place by short-circuited quarter-wave stubs or suitably designed dielectric insulators.

The second form is fabricated by a photochemical process on dielectric substrate sheets ($\varepsilon_r = 2$ to 7, $\tan\delta = 10^{-4}$ to 10^{-3}, and the substrate thickness is 1.0 to 4 mm) foil-cladded from both sides. The foil thickness is 15 to 100 μm which is several times the skin depth. In the unbalanced design, one layer of foil serves as a shield, or ground plane, and the other is etched to form the *line pattern*. In the balanced design, the etched line pattern is covered by a *mating board* fully stripped of its foil on the inner side. The outer 'ground planes' in a strip line are reliably connected together around the periphery of the device and also at several points between the line pattern.

As a way of avoiding dielectric losses, a *high-Q strip line* is used (Fig. 1.5*c*). The central conductor is formed by the parallel-connected foil strips on both sides of a thin dielectric substrate. The substrate is set up on disc supports sandwiched between the metal ground

planes. Practically no electric field exists inside the dielectric substrate, so the dielectric losses have almost no effect on the line parameters.

The third form, or *microstrip lines*, are most exclusively built on dielectric substrates with a high dielectric constant ($\varepsilon_r = 10$ to 15, $\tan\delta = 10^{-4}$ to 5×10^{-4}, with a thickness of 0.5 to 1.0 mm). Their size is about $1/\sqrt{\varepsilon_r}$ that of air-filled rigid strip lines, so they are well adapted to applications involving hybrid integrated circuits.

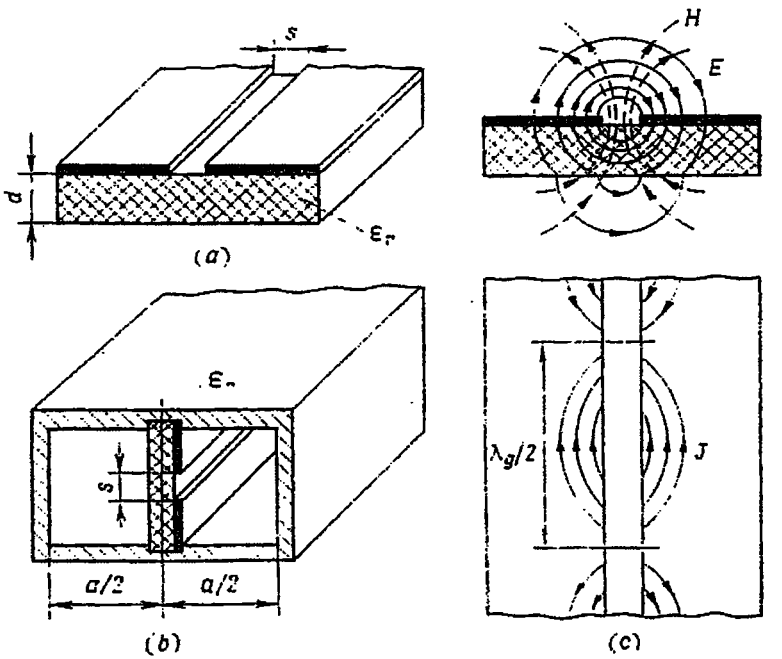


Fig. 1.6. Slot microstrip line

The base of a microstrip line has standard dimensions, with its surface worked to a high finish and polished. The desired line pattern is applied to a substrate by vacuum sputtering through a mask, with the conductors then built up to the required thickness by electroplating. The most commonly used form of microstrip line is shown in Fig. 1.5*a*.

In unbalanced dielectric-filled strip lines the lowest mode has a quasi-TEM structure. Its phase velocity stands mid-way between the velocities of light in the dielectric and air.

The characteristic impedance of strip and microstrip lines ranges between 20 and 100 ohms and can readily be adjusted by varying the conductor width. It is a formidable task to calculate the parame-

ters of strip and microstrip lines by hand; so this is done on computers, using suitable programs. For some strip and microstrip lines, the key parameters can be found in reference sources [9].

Strip and microstrip line sections set up on a common base may affect one another. As has been found, the interaction of parallel lines is negligible if the spacing between the conductors is twice or three times their width. With a closer spacing, distributed electromagnetic coupling begins to be felt; it is utilized by directional couplers fabricated in strip line (see Sec. 4.4).

Also classed with strip lines are slotted and coplanar lines. Among their advantages are the possibility of connecting lumped-constant (discrete) devices in both series and shunt, ease of joining to other types of lines, and usability as components for nonreciprocal ferrite devices.

A *slot microstrip line* has a narrow slot in the conductor applied to one side of a dielectric board or sheet which has a high permittivity (Fig. 1.6*a*). This type of line sustains a slowed-down TE wave with its electromagnetic field concentrated near the slot. The cutoff frequency for this dominant mode is zero but the dispersion is noticeable. When a slotted microstrip is enclosed in a rectangular shield, the combination makes what is known as a *fin-line* (Fig. 1.6*b*). It can readily be combined with rectangular waveguide systems and is often used as a component of waveguide-strip radiators.

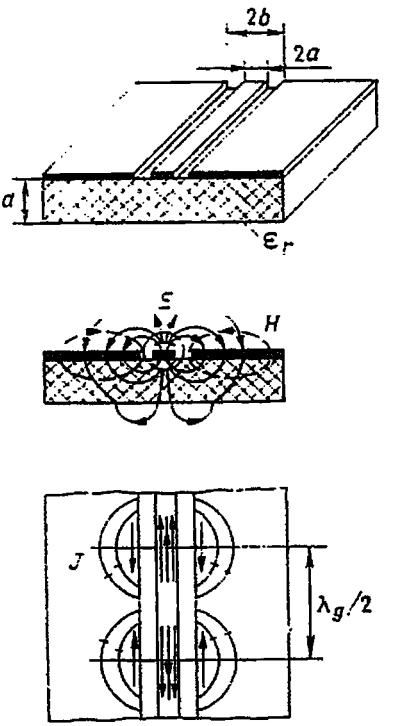


Fig. 1.7. Coplanar microstrip line

A *coplanar microstrip line* (Fig. 1.7) is a three-conductor structure formed by two parallel, closely spaced narrow slots in a metal sheet on one side of a dielectric substrate. As in a slot line, the substrate of a coplanar line has a high permittivity ($\varepsilon_r \geqslant 10$), and this fact is responsible for a marked reduction in the wavelength and the concentration of the field near the central strip. The dominant mode that can be sustained by a coplanar line is the slowed-down TE mode which is subject to a good deal of dispersion but its cutoff frequency is zero. The electric parameters of slotted and coplanar lines are usually calculated on a computer.

Rectangular and circular waveguides. Rectangular metal wave-

guides are used in the decimetric (UHF), centimetric (SHF) and millimetric (EHF) bands to transmit microwave power in the principal TE_{10} (or H_{10}) mode. The cross sectional dimensions, $a \times b$, of a waveguide are chosen so as to strike a balance between the conflicting requirements for maximum power transfer, minimum attenuation, and maximum bandwidth. As the operating wavelength λ_0 approaches the cutoff wavelength, λ_c, $TE_{10} = 2a$, the electric strength of the waveguide rapidly decreases in direct proportion to the factor $[1 - (\lambda_0/2a)^2]^{1/2}$, and the signal attenuation increases rather rapidly in inverse proportion to the same factor. Because of this it is customary to assume that the maximum usable wavelength λ_{max} for rectangular waveguides should be 10% below the cutoff wavelength for the dominant TE_{10} mode.

The minimum wavelength for a rectangular waveguide is chosen so as to prevent the excitation and propagation of higher-order modes of which the TE_{20}, TM_{11} and TE_{11} modes have the lowest cutoff wavelength. For the TE_{20} mode it is equal to a, whereas for the TM_{11} and TE_{11} modes it is equal to $2ab/(a^2 + b^2)^{1/2}$. To be fully sure that no higher-order modes exist in the waveguide, the minimum operating wavelength λ_{min} should be at least 1% above the cutoff wavelength for the first higher-order mode.

The cross sectional dimensions, $a \times b$, for rectangular waveguides are standardized. The parameters of some standard copper waveguides of Soviet manufacture are summarized in Table 1.1. The maxi-

Table 1.1

$a \times b$, mm	λ_c, cm	λ_{av}, cm	λ_{min}-λ_{max}, cm	α_m, dB/m	P_s, kW
72×34	14.4	10	7.7-13	0.02	3000
28.5×12.5	5.7	3.2	2.9-5.1	0.07	500
23×10	4.6	3.2	2.3-4.1	0.12	300
10.6×4.3	2.12	1.25	1.07-1.9	0.35	70
7.2×3.4	1.44	0.8	0.75-1.2	0.51	40

mum safe power, P_s, quoted in the table applies to air-filled waveguides under normal atmospheric conditions. The power level can be raised several-fold by filling the waveguide system with SF_6 gas and maintaining a positive pressure of 3 to 5 $\times$ 10^5 Pa inside it.

The cross-section of a rectangular waveguide can be reduced if it is fabricated in the form of an H- or a Π-shape respectively known as the *double-ridged waveguide* and the *single-ridged waveguide* (Fig. 1.8). The cutoff wavelength for the highest TE modes in these waveguides may be increased without bound if we decrease the gap g where the

transverse electric field is concentrated. Note that the fin-line mentioned earlier may also be treated as a double-ridged waveguide using an inhomogeneous dielectric. The cutoff wavelengths for the highest TE modes in double- and single-ridged waveguides rise only slightly, and this serves to expand their bandwidth to several octaves.

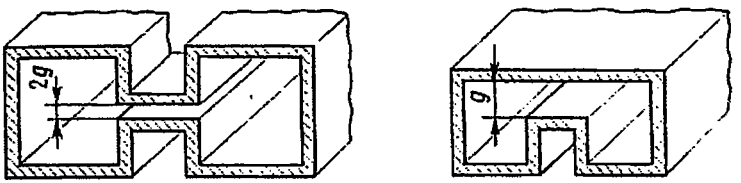


Fig. 1.8. Rigid waveguides: *a*, double-ridged; *b*, single-ridged

However, the increase in the bandwidth of ridged waveguides is accompanied by an impairment in their electric strength and a marked increase in the attenuation constant.

Circular waveguides are mainly used to make components for microwave devices and also, but more seldom, to convey microwave

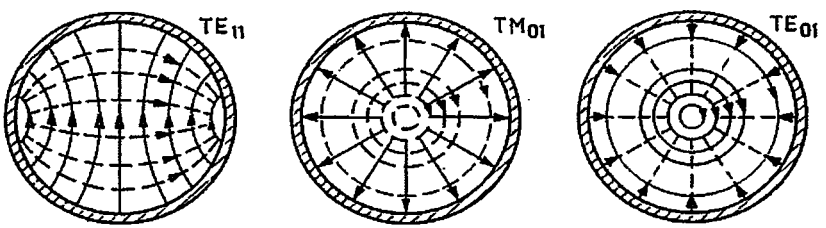


Fig. 1.9. Field patterns in the cross-section of a circular waveguide

energy over considerable distances. The first six modes that can be excited and propagated in a circular waveguide of radius a are listed in descending order of cutoff wavelength in Table 1.2.

Table 1.2

Mode	TE_{11}	TM_{01}	TE_{21}	TM_{11}	TE_{01}	TE_{31}
λ_c/a	3.42	2.61	2.06	1.64	1.64	1.46

The most commonly used modes are TE_{11}, TE_{01} and TM_{01}. Their field patterns over the cross-section of a waveguide are shown in Fig. 1.9.

In terms of both field pattern and properties, the dominant TE_{11} mode is analogous to the TE_{10} mode supported by a rectangular waveguide. However, a circular waveguide may additionally sustain two degenerate TE_{11} modes (that is, those having the same cutoff frequency) with orthogonally polarized fields, say, horizontal and vertical. In an ideal circular waveguide these modes are theoretically independent of each other. However, any discontinuity in a real waveguide may cause transfer of energy from, say, the horizontally to the vertically polarized mode, that is, a change in polarization as the modes are propagated down the waveguide. The linear polarization of the TE_{11} mode can be stabilized by resort to an oval or an elliptical waveguide. In order that such waveguides could be bent as may be required when installing long waveguide circuits, the metal walls of the pipes are made of corrugated metal foil protected by a dielectric coating from the outside. Automatic units have been developed to fabricate such semirigid oval waveguides.

The axially symmetrical TM_{01} mode excited in a circular waveguide is employed in rotary (rotating) joints, but measures need to be taken to prevent the excitation of the TE_{11} mode which has a greater cutoff wavelength, λ_c.

The axially symmetrical TE_{01} mode excited in a circular waveguide has a relatively low cutoff wavelength and, as follows from Table 1.2, it can be accompanied by at least four more modes.

A distinction of the TE_{01} mode in a circular waveguide is the absence of longitudinal electric current components at the walls and a fast decrease in the amplitude of transverse current components as waveguide diameter increases or operating wavelength decreases. Therefore, the attenuation of the TE_{01} mode in a circular waveguide asymptotically tends to zero as the ratio of waveguide radius to wavelength is increased. Because of this, TE_{01} circular waveguides may be used for long-distance transmission of millimetric waves and to build high-Q cavity resonators.

Surface-wave transmission lines. One way to reduce losses in transmission lines is to use slowed-down surface waves. A good proportion of the energy carried by a surface wave is transferred in free space above a *slow-wave structure*, and this brings down the attenuation. A property common to all slowed-down waves is the fact that the amplitude of the field rapidly decreases (practically in an exponential manner) on moving away from the slow-wave structure. The concentration of a surface wave is customarily stated in terms of the *field radius* describing around the slow-wave structure a circle containing 99% of the wave power. As the field radius is increased, the coupling between the surface wave and the slow-wave structure grows weaker, and the attenuation is reduced. Unfortunately, when the field radius is very great, the surface wave becomes unstable and can readily be emitted from irregularities, especially from line

bends. Practically, the field radius ought not to exceed several wavelengths.

The most commonly used form of surface-wave transmission lines is the *dielectric-rod line.** As its name implies, this line is a solid or a hollow dielectric rod along which an axially nonsymmetric hybrid mode is propagated (Fig. 1.10). This mode is excited by gradually deforming the TE_{11} mode of a circular waveguide via a horn launcher. In fact, two such modes orthogonally polarized can exist

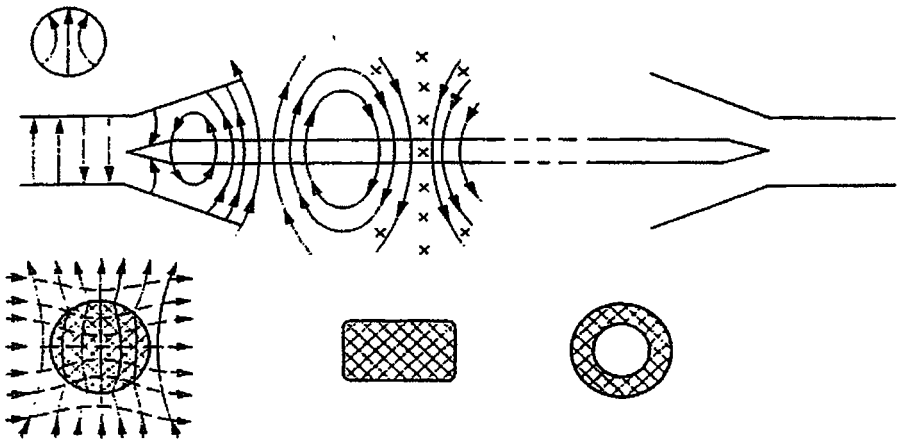


Fig. 1.10. Dielectric-rod transmission line

in a dielectric-rod waveguide. Circular dielectric rods are used but seldom because the orthogonally polarized modes on them have the same phase velocity and strongly interact at discontinuities. Preference is therefore given to rectangular and elliptical dielectric rods.

The cutoff frequency for the dominant hybrid mode on a dielectric-rod waveguide is, theoretically, equal to zero. However, already at below what is known as the *real cutoff frequency* the slow-down ratio decreases so much and the field radius increases to such an extent that the wave becomes unstable and decays owing to a heavy emission. Arbitrarily, the real cutoff frequency is taken that for which the *normalized relative phase velocity*, $n_N = c/v_p - 1$, is about 10^{-3} and the field radius is about $10\,\lambda_0$, or ten times the free-space wavelength.

With the normalized relative phase velocity, n_N, ranging in value from 10^{-2} to 10^{-3}, the attenuation of a surface wave is from one-hundredth to one-tenth of the value existing in the case of an unbounded dielectric medium. At millimetric wavelengths (the EHF

* Some authors place dielectric-rod lines (or waveguides) in a class of its own and reserve the name 'surface-wave' only for single-conductor lines.—*Translator's note.*

band), the attenuation is 0.03 to 0.1 dB/m, which is by an order of magnitude lower than it is in a copper rectangular waveguide. With the normalized relative phase velocity ranging in value from 0.1 to 0.3, the attenuation in a dielectric-rod line is practically the same as that existing in an unbounded dielectric medium. The field radius is about twice the rod radius, and the line may be bent to a radius of ten wavelengths in free space, $10\lambda_0$.

Another variety of surface-wave lines, known as *image lines*, are combinations of dielectric rods and metal sheets, arranged as

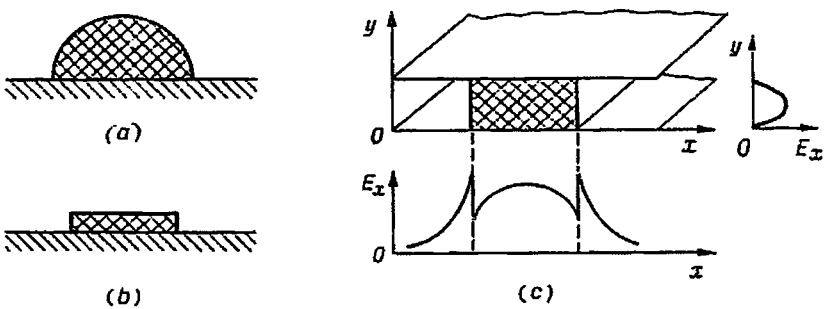


Fig. 1.11. Dielectric waveguides for millimetric wavelengths

shown in Fig. 1.11*a* and *b*. The field patterns in such lines, including the field in the image planes, are the same as in conventional dielectric-rod waveguides, but the metal image plane stabilizes the polarization of the wave. At frequencies around 70 GHz image lines have an attenuation of 0.02 to 0.2 dB/m.

Figure 1.11*c* shows a rectangular dielectric rod sandwiched between two parallel metal image planes. The resultant slowed-down hybrid wave does not give rise to any longitudinal electric-field components, so the losses in the image planes decrease with increasing frequency. The amplitude of the field set up by this wave decreases exponentially at a high rate on moving away from the dielectric rod. Owing to this, the image planes may be made small and the line itself more compact.

Optical-fibre transmission lines. At submillimetric and optical wavelengths conventional transmission lines cannot be used because the losses become prohibitively heavy and the lines themselves are difficult to make. These limitations have led to the use of *optical-fibre transmission lines.*

An optical-fibre transmission line is a multi-mode waveguide made in the form of bundles of fine fibres about 150 μm in diameter (which is equivalent to several hundred wavelengths in the optical region) fabricated from ultra-pure quartz which introduces negligible losses at optical wavelengths.

For its operation, an optical-fibre line depends on the propagation of light energy inside a quartz fibre by total internal reflection. To avoid losses on surface irregularities, the total-internal-reflection boundary should be shifted inside the fibre. This objective is achieved by causing the optical coefficient of refraction to vary stepwise or continuously in the radial direction. The desired radial profile of the refractive index, n, is obtained by doping the host material with the oxides of boron, phosphorus and some other elements. As the frequency of the incident wave is varied, the path travelled by the

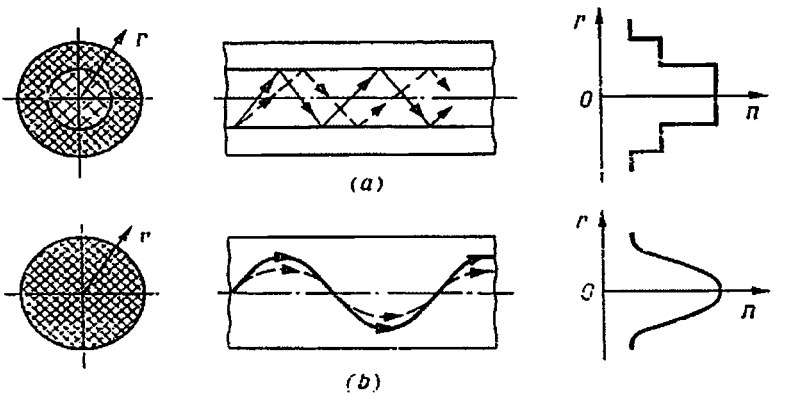


Fig. 1.12. Optical-fibre waveguides: *a*, step-index profile; *b*, graded-index profile

rays inside the optical fibre changes as shown in Fig. 1.12. This gives rise to dispersion and distortion in the transmission of broadband signals. Dispersion is less noticeable in optical fibres with a graded radial refractive index profile and can be reduced still more by giving this profile the desired shape. Optical fibres are given a coat of varnish to protect them against mechanical and environmental factors and bundled to form an optical-fibre (or fibre-optic) cable.

Attenuation in optical-fibre lines is solely due to dielectric losses. State-of-the-art technology is capable of making quartz fibres of exceptionally high transparency and of controlling the percentage of dopants to extremely tight tolerances so as to produce the desired radial refractive index profile. In the most commonly used grades of optical-fibre lines (at this writing) the attenuation is a mere 3 to 5 dB/km.

Excitation sources (drivers) for optical-fibre lines are light-emitting diodes (LEDs) and semiconductor lasers. At the receiving end of optical-fibre lines, signals are picked up by *PIN* diode detectors remarkable for high speed of response and sensitivity or avalanche photodiode (APD) devices capable of signal amplification.

Optical-fibre lines provide very broad bandwidths, ideal noise immunity, and a practically complete isolation (decoupling) between channels. Work is under way on components for optical-fibre lines, such as multi-channel connectors, directional couplers, commutators, filters, and so on. Ever more ground is being gained by single-mode optical-fibre lines, that is, lines operating in only one mode. They are made of optical fibres 3 to 5 μm in diameter, practically free from dispersion.

1.3. The Mathematical Model of a Uniform Transmission Line

Within uniform portions of a transmission line, the electromagnetic field due to a given mode is the superposition of an incident (direct or forward) and a reflected (reverse or backward) wave. The term '*incident*' refers to a travelling wave produced by a generator and propagated from the selected origin, $\xi = 0$, in the positive ξ-direction. Assuming that the time dependence of monochromatic components is of the exp ($j\omega t$) form, any component of the electromagnetic field associated with the incident wave is connected to the longitudinal coordinate ξ by a relation of the exp ($-\dot{\gamma}\xi$) form, where

$$\dot{\gamma} = \beta - j\alpha$$

is the complex propagation constant, α is the attenuation constant in nepers per metre of line length, and β is the phase constant in m^{-1}. In turn the phase constant is connected to the guide wavelength λ_g and phase velocity v_p by a well-known relation

$$\beta = 2\pi/\lambda_g = \omega/v_p$$

Recall that the guide wavelength is defined as the distance over which the phase of a travelling wave changes by 2π.

The term '*reflected*' refers to a travelling wave produced either by a line discontinuity or load and propagated in the direction opposite to the incident wave. Any field component of a reflected monochromatic wave on a uniform transmission line is connected to the longitudinal coordinate by a relation of the exp ($+\dot{\gamma}\xi$) form.

The incident and the reflected electromagnetic waves transfer each a certain amount of electromagnetic energy along the transmission line. It is established in electrodynamics that the incident and the reflected wave are *orthogonal* to each other. This means that the active power conveyed by any of the two waves is independent of the other, propagated in the opposite direction. The active power carried by the incident wave is the integral of the Poynting vector of the wave, taken over the cross-section of the transmission line [see

Eq. (A.7) in the Appendix]:

$$P_i = Re\left\{\int_S [\dot{\mathbf{E}}_i,\ \dot{\mathbf{H}}_i^*]\, d\mathbf{S}\right\} \tag{1.4}$$

where $\dot{\mathbf{E}}_i$ and $\dot{\mathbf{H}}_i$ are the vector distribution functions of the rms electric and magnetic fields associated with the incident wave at a given section of the transmission line, and $d\mathbf{S} = \mathbf{n}dS$ is a vector element of the cross-sectional area S.

Open-wire transmission lines have an infinite cross-sectional area, so for them the integral in Eq. (1.4) becomes improper. However, in an open-wire transmission line the electric and magnetic field strength diminishes rapidly on moving away from the conductors and dielectrics, so the integral in (1.4) converges and has a finite value.

The power conveyed by a reflected wave, P_r, is given by an equation similar to Eq. (1.4), except that $\dot{\mathbf{E}}_i$ and $\dot{\mathbf{H}}_i$ are replaced by $\dot{\mathbf{E}}_r$ and $\dot{\mathbf{H}}_r$, differing solely in the proportionality factor.

The description of the electromagnetic field in a transmission line in terms of the vector functions $\dot{\mathbf{E}}(v_1, v_2, \xi)$ and $\dot{\mathbf{H}}(v_1, v_2, \xi)$ (where v_1 and v_2 are coordinates in the plane of the cross-section) carries much more information than may be necessary for the analysis and design of a transmission system. Indeed, a transmission system designer is mainly interested in the amount of power to be transmitted, the relative magnitude of the incident and reflected waves, the phase delay and attenuation over a definite length of the transmission line. But these parameters (power, transmission gain, and reflection coefficient) can quite readily be found by experiment, whereas the measurement of the field components and their distribution functions run into appreciable difficulties. Moreover, once a particular type of transmission line has been selected a detailed knowledge of the field patterns, vectors distribution and directions are of secondary importance, and these factors should preferably be excluded from consideration in design.

This objective is achieved by building a universal mathematical model in the form of an *equivalent long line*. In an equivalent long line, the vector field distribution functions at cross-sections are replaced by integral (averaged) measures of the electromagnetic field — the *equivalent normalized incident* and *reflected voltage waves*. These waves are defined as follows:

$$\dot{v}_i(\xi) = (P_i)^{1/2} \exp(j\varphi_i^E) \exp(-\dot{\gamma}\xi) = \dot{v}_i(0) \exp(-\dot{\gamma}\xi)$$

$$\dot{v}_r(\xi) = (P_r)^{1/2} \exp(j\varphi_r^E) \exp(+\dot{\gamma}\xi) = \dot{v}_r(0) \exp(+\dot{\gamma}\xi) \tag{1.5}$$

As follows from Eqs. (1.5), the magnitude of the normalized (incident or reflected) voltage wave is the square root of the respective

power and has therefore the dimensions of $W^{1/2}$. The phase of the respective normalized voltage wave is taken to be equal to that of the transverse component of the electric field associated with the respective wave. For better insight into the concept of normalized travelling voltage wave, let us refer to examples.

Example 1.1. A coaxial waveguide operating in the TEM mode (Fig. 1.13). The cutoff frequency, ω_c, for this mode is zero. At any plane cutting across the waveguide, only the field components $\dot{E}_r$ and $\dot{H}_\varphi$ exist, having the same radial distribution, $1/r$. The ratio of the transverse components $\dot{E}_r$ and $\dot{H}_\varphi$ gives the *wave impedance*

$$Z_w^{\text{TEM}} = \dot{E}_r/\dot{H}_\varphi$$

Whatever the configuration of the conductors in the case of the TEM mode, the characteristic wave or, simply, wave impedance Z^{TEM} is

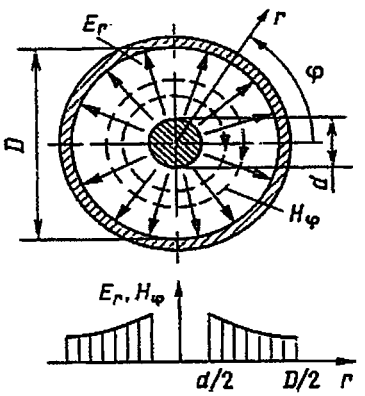


Fig. 1.13. TEM-mode coaxial waveguide

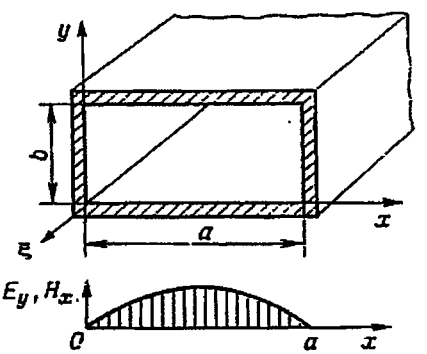


Fig. 1.14. Rectangular waveguide operating in the TE_{10} mode

solely determined by the parameters of the magnetodielectric material which fills the line, that is,

$$Z_w^{\text{TEM}} = (\mu_a/\varepsilon_a)^{1/2}$$

For a travelling TEM mode in a coaxial waveguide, the magnitude of the Poyinting vector is

$$\Pi_t(r) = (E_{\max}^2/Z_w^{\text{TEM}})\,(d^2/4r^2)$$

where $E_{\max}$ is the rms value of the radial component of the electric field near the surface of the inner conductor with $r = d/2$ (the subscript 'max' indicates that this is the maximum value within the

cross-section of the coaxial waveguide). Using Eq. (1.4), we obtain

$$\dot{v}_i(0) = (\dot{E}_{\max}/\sqrt{Z_w^{\text{TEM}}})(d/2)\left(\int_0^{2\pi}\int_{d/2}^{D/2}\frac{dr\,d\varphi}{r}\right)^{1/2}$$

$$= \dot{E}_{\max} d\left[\frac{\pi \ln(D/d)}{2Z_w^{\text{TEM}}}\right]^{1/2} \qquad (1.6)$$

Example 1.2. A rectangular waveguide operating in the TE_{10} mode (Fig. 1.14). The electromagnetic field in this mode has three components, $\dot{E}_y$, $\dot{H}_x$ and $\dot{H}_z$. The transverse components $\dot{E}_y$ and $\dot{H}_x$ have an identical distribution over the cross-section, $\sin(\pi x/a)$. Note that for an incident TE-wave in any hollow waveguide the ratio between $\dot{E}_t$ and $\dot{H}_t$ is the same at any point within the cross-section and equal to the *wave impedance* for that mode:

$$Z_w^{\text{TE}} = (\mu_a/\varepsilon_a)^{1/2}(\lambda_g/\lambda_0) = (\mu_a/\varepsilon_a)^{1/2}[1 - (\lambda_0/\lambda_c)^2]^{-1/2}$$

where ε_a = permittivity of the dielectric filling the waveguide
μ_a = permeability of the dielectric filling the waveguide
λ_0 = launcher wavelength
λ_c = cutoff wavelength, considering the dielectric that fills the waveguide

The magnitude of the Poynting vector for the traveling TE_{10} mode is given by

$$\Pi_t = (E_{\max}^2/Z_w^{\text{TE}})\sin^2(\pi x/a)$$

where $E_{\max}$ is the rms value of the electric field strength at the axis of the waveguide, that is, at antinode of the cross-sectional distribution. Using Eq. (1.4), we obtain

$$\dot{v}_i(0) = (\dot{E}_{\max}/\sqrt{Z_w^{\text{TE}}})\left[\int_0^b\int_0^a \sin^2(\pi x/a)\,dx\,dy\right]^{1/2} = \dot{E}_{\max}(ab/2Z_w^{\text{TE}})^{1/2}$$

Reflection coefficient. The ratio between the transverse electric field components for an incident and a reflected wave at the same point of the cross-section of a transmission line is called the *electric-field reflection coefficient* of that line

$$\dot{\rho}_E = \dot{E}_{t,r}/\dot{E}_{t,i}$$

Since, according to Eq. (1.5), the normalized incident and reflected voltage waves are practically proportional to the transverse components of the respective electric fields and have the same phases, it follows that the reflection coefficient for the normalized voltages,

$$\dot{\rho} = \dot{v}_r/\dot{v}_i$$

is always identical with the reflection coefficient for the electric field, $\dot{\rho}_E$. Noting how the incident and reflected voltage waves vary with distance along the waveguide axis, it is an easy matter to establish how the reflection coefficient varies from one transverse plane to another:

$$\dot{\rho}(\xi) = \dot{v}_r(\xi)/\dot{v}_i(\xi) = \dot{v}_r(0)\exp(j\dot{\gamma}\xi)/\dot{v}_i(0)\exp(-j\dot{\gamma}\xi)$$
$$= \dot{\rho}(0)\exp(2j\dot{\gamma}\xi) \qquad (1.7)$$

where $\dot{\rho}(0)$ is the reflection coefficient at section $\xi = 0$, and $\dot{\gamma} = \beta - j\alpha$ is the propagation constant for the mode in question.

Note that along with the ξ-coordinate directed from generator to load, use is often made of the l-coordinate pointing in the opposite direction (Fig. 1.15). Then the reflection coefficient varies with distance along the axis as

$$\dot{\rho}(l) = \dot{\rho}(0)\exp(-2j\dot{\gamma}\xi)$$

No matter what equation is used to express the reflection coefficient, when the transverse plane in question moves towards the genera-

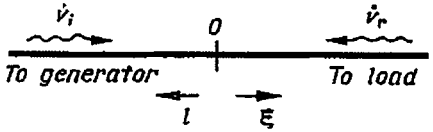


Fig. 1.15. Waves in an equivalent long line

tor the reflection coefficient suffers a phase lag and decreases in absolute value owing to the attenuation in the transmission line.

Total normalized voltages and currents on a transmission line. When an incident and a reflected wave coexist on a transmission line, the active power transmitted through a selected transverse plane is equal to the difference between the powers transferred by the incident and reflected waves:

$$P = P_i - P_r = |\dot{v}_i|^2 - |\dot{v}_r|^2 = |\dot{v}_i|^2(1 - |\dot{\rho}|^2) \qquad (1.8)$$

The transmitted active power may be re-written in a different form

$$P = \mathrm{Re}\{[\dot{v}_i(1+\dot{\rho})][\dot{v}_i^*(1-\dot{\rho}^*)]\} = |\dot{v}_i|^2\,\mathrm{Re}[(1-|\dot{\rho}|^2) + (\dot{\rho}-\dot{\rho}^*)]$$

which is legitimate to do because the difference $(\dot{\rho} - \dot{\rho}^*)$ is a pure imaginary (the asterisk * denotes the complex conjugate). Therefore, we may, at least formally, introduce the following definitions for any mode propagated on a transmission line:

(1) The total normalized voltage (with dimensions of $W^{1/2}$) is the sum of the normalized incident and reflected voltage waves:

$$\dot{v} = \dot{v}_i + \dot{v}_r = \dot{v}_i (1 + \dot{\rho}) \tag{1.9}$$

(2) The total normalized current (with dimensions of $W^{1/2}$) is the difference between the normalized incident and reflected voltage waves:

$$\dot{i} = \dot{v}_i - \dot{v}_r = \dot{v}_i (1 - \dot{\rho}) \tag{1.10}$$

As follows from Eq. (1.9), the normalized voltage on a transmission line is proportional to the total electric field which exists on the line owing to the superposition of the incident and reflected waves.

To obtain insight into the concept of normalized current as defined by Eq. (1.10), it should at first be noted that the negative reflection coefficient is nothing but the reflection coefficient for the magnetic field

$$-\dot{\rho} = \dot{\rho}_H = \dot{H}_{t,r}/\dot{H}_{t,i}$$

(Because $\dot{\rho}_E$ and $\dot{\rho}_H$ are in antiphase, the Poynting vectors for the incident and reflected waves are opposite in orientation.) Therefore, the normalized current defined by Eq. (1.10) may be written as

$$\dot{i} = \dot{v}_i (1 + \dot{\rho}_H)$$

and turns out to correspond to the total magnetic field produced by the superposition of the incident and reflected waves. This is the reason why it is legitimate to use the term 'current'. In view of the above definitions for the total normalized voltage and current, the equation describing the active power transmitted through the selected transverse plane takes the same form as is used in the theory of low-frequency electric circuits:

$$P = \mathrm{Re}\,(\dot{v}\dot{i}^*) \tag{1.11}$$

Normalized impedance and admittance. The ratio between the total normalized voltage $\dot{v}$ on an equivalent long line and the total normalized current $\dot{i}$ in the same line is termed the *normalized transmission-line impedance* (which is a dimensionless quantity):

$$\dot{z} = r + jx = \dot{v}/\dot{i} = (1 + \dot{\rho})/(1 - \dot{\rho}) \tag{1.12}$$

Its reciprocal is termed the *normalized transmission-line admittance* (which is likewise a dimensionless quantity):

$$\dot{y} = g + jb = (1 - \dot{\rho})/(1 + \dot{\rho}) = 1/\dot{z}$$

Using the concept of normalized transmission-line impedance and admittance, the power transferred by a line may be presented in two

more forms employed in the theory of low-frequency circuits:

$$P = |\dot{v}|^2 g = |\dot{i}|^2 r$$

It follows from Eq. (1.12) that when the reflection coefficient is zero, that is, when only an incident wave exists, the ratio of the normalized voltage to the normalized current is equal to unity. This implies that a *dimensionless characteristic impedance* of unity is automatically set in the equivalent long line. It should be stressed once more that the total normalized voltage on an equivalent long line is defined as the sum of the incident and reflected voltage waves, and the total normalized current as their difference. There is no other meaning in the concept of normalized voltage and current. In the general case, there is even no way of indicating the points in the cross-section of a line between which the normalized voltage exists and which conductor carries the normalized current.

Normalization. For transmission lines operating in TEM modes (and only in such modes!), we are in a position to introduce the rigorous concepts of an incident voltage wave $\dot{V}_i$ expressed in volts, an incident current wave $\dot{I}_i$ expressed in amperes, and the characteristic impedance $Z_c = \dot{V}_i/\dot{I}_i$, expressed in ohms. Using these non-normalized quantities, the conditions existing on a uniform section (or segment) of a transmission line operating in the TEM mode can be described by relations known from what may be called the conventional theory of long lines

$$\dot{V} = \dot{V}_i + \dot{V}_r = \dot{V}_i(1 + \dot{\rho}), \quad \dot{I} = \dot{I}_i + \dot{I}_r = (\dot{V}_i/Z_c)(1 - \dot{\rho}),$$
$$\dot{Z} = R + jX = \dot{V}/\dot{I} = Z_c[(1 + \dot{\rho})/(1 - \dot{\rho})] = 1/(G + jB) \quad (1.13)$$

The power transferred on a long line is given by

$$P = \frac{|\dot{V}_i|^2}{Z_c}(1 - |\dot{\rho}|^2) = \mathrm{Re}(\dot{V}\dot{I}^*) = |\dot{V}|^2 G = |\dot{I}|^2 R \quad (1.14)$$

In the above equations the reflection coefficient $\dot{\rho}$ has the same meaning as in Eqs. (1.9) and (1.10) for an equivalent long line.

By comparing Eqs. (1.13) and (1.14) with Eqs. (1.8) and (1.11), we can write the following normalizing relations for voltages and currents on a transmission line operating in the TEM mode:

$$\begin{aligned} \dot{v}_i &= \dot{V}/\sqrt{Z_c} = \dot{I}_i\sqrt{Z_c} \\ \dot{v} &= \dot{V}/\sqrt{Z_c} \\ \dot{i} &= \dot{I}\sqrt{Z_c} \end{aligned} \quad (1.15)$$

In turn, Eqs. (1.15) may be used to write the rule for the normalization of impedances:

$$\dot{z} = \dot{v}/\dot{i} = \dot{V}/(\sqrt{Z_c}\,\dot{I}\,\sqrt{Z_c}) = \dot{Z}/\dot{Z}_c = (1+\dot{\rho})/(1-\dot{\rho}) \qquad (1.16)$$

Example 1.3. Normalize the incident voltage and current in a coaxial waveguide. In a coaxial waveguide operating in the TEM mode (Fig. 1.13), the voltage $\dot{V}_i$ between the inner and outer conductors can be found by integrating the radial component of the electric field intensity:

$$\dot{V}_i = \int_{d/2}^{D/2} \dot{E}_r\,dr = \int_{d/2}^{D/2} (E_{max}d/2r)\,dr = (\dot{E}_{max}d/2)\ln(D/d)$$

where $\dot{E}_{max}$ is the radial electric field strength or intensity (in V/m) on the surface of the inner conductor at $r = d/2$.

By Ampere's circuital law, the longitudinal current $\dot{I}_i$ flowing on the surface of the inner conductor is given by the integral of the tangential component of the magnetic field:

$$\dot{I}_i = \int_0^{2\pi} \dot{H}_\varphi r\,d\varphi|_{r=d/2} = 0.5d\int_0^{2\pi} \dot{H}_{max}\,d\varphi$$

where

$$\dot{H}_{max} = E_{max}/Z_w^{TEM}$$

is the azimuthal magnetic field strength or intensity (in A/m) at $r = d/2$. On evaluating the integral, we obtain

$$\dot{I}_i = \dot{E}_{max}\pi d/Z_w^{TEM}$$

The ratio of the voltage to the current associated with a travelling wave gives the characteristic impedance (in ohms) of the coaxial waveguide in question:

$$\dot{Z}_c = \dot{V}_i/\dot{I}_i = (Z_w^{TEM}/2\pi)\ln(D/d) = (\mu_r/\varepsilon_r)^{1/2} \times 60\ln(D/d)$$

On inserting the values of $\dot{V}_i$, $\dot{I}_i$ and Z_c thus found in Eqs. (1.15), we obtain the normalized incident voltage:

$$\dot{v}_i = \dot{V}_i/\sqrt{Z_c} = \dot{I}_i\sqrt{Z_c} = \dot{E}_{max}d\,[\pi\ln(D/d)/2Z_w^{TEM}]^{1/2}$$

which checks with the result found earlier, Eq. (1.6).

The normalization equations (1.15) and (1.16) may be used in a two-fold manner. On the one hand, they readily lead to the normalized voltages, currents and impedances for transmission lines operating in the TEM mode without recourse to Eqs. (1.5). On the

other (which is less obvious), Eqs. (1.15) and (1.16) offer an opportunity to go over from the mathematical model of a line in the form of an equivalent long line with a dimensionless characteristic impedance of unity to other mathematical models in the form of equivalent long lines having an arbitrary characteristic impedance (usually likewise dimensionless). Resort to non-unity characteristic impedances is justified, for example, when dealing with junctions of two transmission lines differing in cross-sectional area. The reflections from each junction may be ascribed in equivalent circuits to abrupt changes in the dimensionless characteristic impedance of the equivalent long lines converging to the junction. The magnitude of this abrupt change is assumed to be such that the reflection coefficient of the prototype junction would be the same as that of its model in the equivalent circuit.

1.4. The Effect of Mismatch on Transmission Line in Efficiency and Power Transfer

In the general case, an arbitrary load will give rise to a reflected wave on a transmission line. When it is superimposed on the incident wave, the reflected wave produces multiple maxima and minima in

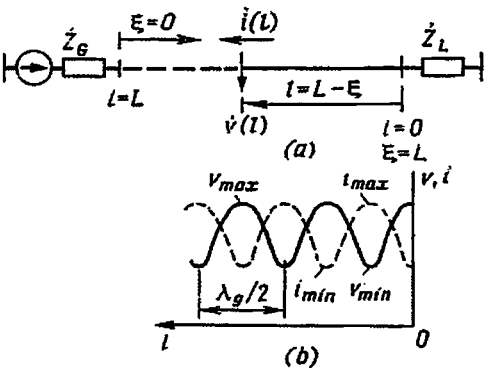


Fig. 1.16. Power transfer from generator to load over a line: *a*, line schematic; *b*, current and voltage distribution

the longitudinal distributions of normalized voltages and currents (Fig. 1.16), so that a *mixed wave pattern* is formed. In engineering practice, it is customary to characterize the mixed wave condition in terms of the *travelling wave ratio*, TWR, symbolized as K in this text and defined as the ratio of the minimum value of the total normalized voltage (current or field strength) to the maximum value of the respective quantity on the line:

$$\text{TWR} = K = v_{\min}/v_{\max} = i_{\min}/i_{\max} = (1 - |\dot{\rho}|)/(1 + |\dot{\rho}|)$$

where $|\dot{\rho}|$ is the magnitude of the reflection coefficient. Frequently, instead of the TWR, use is made of its reciprocal, called the *standing wave ratio*, SWR*. The TWR or the SWR is measured by interposing what is known as a *slotted* (or *measuring*) *line* in the transmission line of interest and exploring its longitudinal field distribution with a movable probe.

In the absence of ohmic losses, the magnitude of the reflection coefficient is independent of the longitudinal coordinate (the position on the line), so all minima and maxima in the longitudinal field distribution on the line are respectively the same, and the TWR remains unchanged all along the line. In a lossy transmission line, the magnitude of the reflection coefficient varies on moving over a distance l from load towards generator as follows:

$$|\dot{\rho}(l)| = |\dot{\rho}_L| \exp(-2\alpha l)$$

where $\dot{\rho}_L$ is the coefficient of reflection from the load at $l = 0$.

Because the magnitude of the reflection coefficient for a lossy line is a varying quantity, the respective TWR should be defined subject to certain conditions, namely as the ratio between an adjacent minimum and maximum in the longitudinal voltage distribution. In a uniform lossy line, the TWR increases on moving away from the load. Therefore, the conditions existing on a lossy transmission line may be characterized in terms of two values of the TWR: at load, K_L, and at the generator, K_G. For a uniform transmission line, the two values are connected by a relation of the form

$$K_G = \frac{(1+K_L)-(1-K_L)\exp(-2\alpha L)}{(1+K_L)+(1-K_L)\exp(-2\alpha L)}$$

which is illustrated in the plots of Fig. 1.17. Using these plots, it is an easy matter to find K_G from the specified values of K_L and αL. Note that the relation for K_G is valid only in the case of a uniform transmission line. Accumulation of reflections from a series of discontinuities may impair the TWR on moving away from a weakly reflecting load.

Power transfer to the load is evaluated in terms of the *efficiency* of the transmission line, defined as the ratio between the power dissipated in the load, P_L, and the incident power delivered by the generator to the line, P_G. As the incident wave travels all of the length L of the line, some of the power carried by the incident wave is lost due to attenuation, so the incident power at the load is equal to $P_I \exp(-2\alpha L)$ where the factor

$$\exp(-2\alpha L) = \eta_\alpha$$

* In fact, only the SWR (especially the voltage standing wave ratio, VSWR) is used outside the Soviet Union.—*Translator's note.*

is called the *power attenuation factor* for the incident wave. The reflection of the incident wave from the load leads to a further decrease in the power delivered to the load by a factor $\eta_r = 1 - |\dot{\rho}_L|^2$, so that the efficiency of the transmission line is

$$\eta = P_L/P_i = \eta_\alpha \eta_r = \exp(-2\alpha L)(1 - |\dot{\rho}_L|^2)$$
$$= 4K_L \exp(-2\alpha L)/(1 + K_L)^2$$

The above equation has been derived, using the relation

$$|\dot{\rho}_L| = (1 - K_L)/(1 + K_L)$$

where K_L is the TWR of the load. Plots of the line efficiency as a function of the load TWR and of the total attenuation of the line,

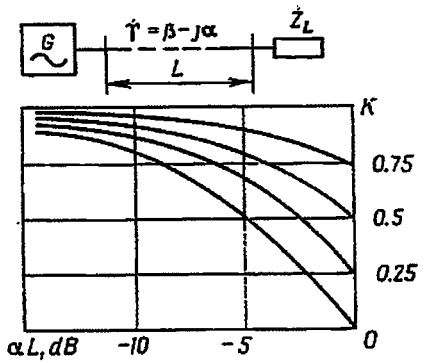


Fig. 1.17. Variations in TWR along a lossy transmission line

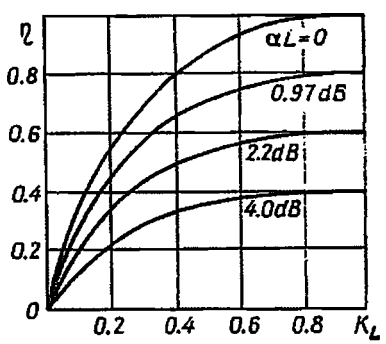


Fig. 1.18. Efficiency of a transmission line as a function of load TWR

αL, in dB, are shown in Fig. 1.18. As is seen, the most favourable conditions for power transfer exist in the case of a matched load, that is, when $K_L = 1$.

Operation of a line into a matched load is likewise favourable from the view-point of the line's electric strength. A breakdown in a mismatched transmission line may occur if the normalized voltage at an antinode in the longitudinal distribution

$$v_{max} = |\dot{v}_i|(1 + |\dot{\rho}|)$$

approaches some critical value, v_{cr}, equal to the square root of the critical power likely to cause a breakdown in the case of a purely travelling wave. On equating v_{max} and v_{cr}, we obtain the condition for a breakdown to occur

$$|\dot{v}_{i,\,cr}|(1 + |\dot{\rho}|) = v_{cr}$$

Squaring both sides and noting that

$$v_{cr}^2 = P_{cr,\,tw}$$

is the breakdown power in the case of a travelling wave, we obtain the incident wave power causing a breakdown in a mismatched waveguide

$$P_{i,\,cr} = |\dot{v}_{i,\,cr}|^2 = P_{cr,\,tw}/(1 + |\dot{\rho}|)^2$$

The above equation shows that a mismatch markedly increases the risk of a breakdown. In a completely mismatched waveguide, a breakdown can be caused by an incident power wave which is a mere quarter of the power causing a breakdown under the travelling-wave conditions.

Sometimes the electric strength of a mismatched waveguide is evaluated, proceeding from the power P_{out} delivered to the load. This power can be expressed in terms of the normalized voltage at an antinode of the longitudinal distribution,

$$v_{max} = |\dot{v}_i|(1 + |\dot{\rho}|)$$

as follows:

$$P_{out} = |\dot{v}_i|^2(1 - |\dot{\rho}|^2) = |\dot{v}_i|^2(1 + |\dot{\rho}|)(1 - |\dot{\rho}|)$$

$$= |\dot{v}_i|^2(1 + |\dot{\rho}|)^2 \frac{1 - |\dot{\rho}|}{1 + |\dot{\rho}|} = v_{max}^2 K$$

where K is the TWR of the transmission line.

On replacing v_{max}^2 by the breakdown power under conditions of a purely travelling wave, $P_{cr,\,tw}$, we find the critical power transferred to the load under conditions of a mismatch:

$$P_{out,\,cr} = P_{cr,\,tw} K$$

Thus, the power that can be transferred to the load just short of a breakdown is a mere fraction of the critical power under conditions of a purely travelling wave, this fraction being proportional to the load TWR.

Consideration of the efficiency and electric strength of a transmission line shows that the most favourable case seems to be when the line and the load are perfectly matched, $K = 1$.

In real microwave transmission (waveguide) systems, a perfect match between the line and the load is never achieved because of many factors (the frequency dependence of the load impedance, additional reflections from discontinuities). Therefore specifications will only stipulate the minimum permissible value of the TWR for both the load and the line as a whole. As a rule, the goal to be achieved is set at a TWR not below 0.7 or 0.8, although in some cases (such as the feeders for receiving HF antennas) the permissible TWR is reduced to 0.3 or 0.4.

1.5. Impedance Transformation in Transmission Lines

Short segments of transmission lines (waveguides) are widely used for load impedance transformation.

The basis for such transformation is the dependence of the normalized impedance at an arbitrary transverse plane of a transmission line on the distance from the load. To derive this dependence, we rewrite Eq. (1.12) as

$$\dot{z}(l) = [1 + \dot{\rho}(l)]/[1 - \dot{\rho}(l)]$$

Impedance transformers are usually built with lines low in losses. This offers the possibility to neglect attenuation and to define the reflection coefficient as an idealized function of distance from the load:

$$\dot{\rho}(l) = \dot{\rho}(0) \exp(-j2\beta l) \tag{1.17}$$

where

$$\beta l = 2\pi l/\lambda_g$$

is the *electrical length* of the line as measured from the load towards the generator.

The complex reflection coefficient at the load, $l = 0$, is connected to the load impedance by the following relations:

$$\dot{z}_L = [1 + \dot{\rho}(0)]/[1 - \dot{\rho}(0)]$$

and

$$\dot{\rho}(0) = (\dot{z}_L - 1)/(\dot{z}_L + 1)$$

On substituting Eq. (1.17) into the expression for $\dot{z}(l)$ and rearranging subject to Euler's equation

$$\exp(\pm j\alpha) = \cos\alpha \pm j \sin\alpha$$

we obtain an equation describing impedance transformation along a transmission line

$$\begin{aligned}\dot{z}(l) &= \frac{1+\dot{\rho}(0)\exp(-j2\beta l)}{1-\dot{\rho}(0)\exp(-j2\beta l)} = \frac{\exp(j\beta l)+\dot{\rho}(0)\exp(-j\beta l)}{\exp(j\beta l)-\dot{\rho}(0)\exp(-j\beta l)} \\ &= \frac{[1+\dot{\rho}(0)]\cos\beta l + j[1-\dot{\rho}(0)]\sin\beta l}{[1-\dot{\rho}(0)]\cos\beta l + j[1+\dot{\rho}(0)]\sin\beta l} \\ &= \frac{\dot{z}_L + j\tan\beta l}{1 + j\dot{z}_L\tan\beta l}\end{aligned} \tag{1.18}$$

In some cases, such as when several loads have to be connected in parallel to the same transmission line, it is convenient to use admit-

tances rather than impedances. On replacing $\dot{z}(l)$ by $1/\dot{y}(l)$ and $\dot{z}_L$ by $1/\dot{y}_L$ in Eq. (1.18), we obtain an equation for admittance transformation

$$\dot{y}(l) = (\dot{y}_L + j \tan \beta l)/(1 + j\dot{y}_L \tan \beta l) \qquad (1.19)$$

which has the same form as the equation for impedance transformation.

Transmission-line (waveguide) segments used for impedance transformation may differ from those used in the transmission (feeder) system proper in cross-sectional area, construction, and even the mode used. In equivalent circuit diagrams, this point is accounted for by introducing non-unity dimensionless characteristic impedances z_c. Equations for impedance and admittance transformation for a transformer having a non-unity characteristic impedance are derived from Eqs. (1.18) and (1.19) by re-normalizing $\dot{z}$ to $\dot{z}/z_c$ and $\dot{y}$ to $\dot{y}z_c$:

$$\dot{z}(l) = z_c \frac{\dot{z}_L + jz_c \tan \beta l}{z_c + j\dot{z}_L \tan \beta l}, \quad \dot{y}(l) = \frac{1}{z_c} \frac{\dot{y}_L z_c + j \tan \beta l}{1 + j\dot{y}_L z_c \tan \beta l} \qquad (1.20)$$

Note that the impedances and admittances in Eq. (1.20) are dimensionless and normalized for the main transmission-line (waveguide) system. The value of z_c in Eq. (1.20) depends on how the impedance transformer(s) differ from the main transmission-line system in cross-sectional geometry (and dielectric parameters). The numerical value of z_c for each specific case is found by electrodynamic calculations or can be looked up in reference sources.

Let us examine several specific types of impedance transformers.

(1) **Reactive stubs.** This term refers to transmission-line segments shorted or open-circuited on the load side. From Eqs. (1.20) we may write for the input reactive impedances and admittances of stubs thus:

$$\dot{z}_{sc} = jz_c \tan \beta l, \quad \dot{y}_{sc} = (-j/z_c) \cot \beta l$$
$$\dot{z}_{oc} = -jz_c \cot \beta l, \quad \dot{y}_{oc} = (j/z_c) \tan \beta l$$

Shorted transmission-line segments less than a half-wave length long are often used as matching elements and in distributed-parameter resonant (tuned) circuits. Open-circuited transmission-line segments are used more seldom. Indeed, in hollow waveguides and many other types of transmission lines, the open-circuit condition is undesirable as it might lead to the emission of radiation from the output opening.

(2) **A quarterwave transformer.** This term applies to transmission-line segments with an electrical length equal to $\beta l = \pi/2$. With a length such as that, it follows from Eq. (1.18) and Eq. (1.19)

that

$$\dot{z}\,(\lambda_g/4)\,\dot{z}\,(0) = 1$$

and

$$\dot{y}\,(\lambda_g/4)\,\dot{y}\,(0) = 1$$

This implies that a quarter-wave transmission-line segment transforms the load impedance $\dot{z}_L$ into $1/\dot{z}_L$ numerically equal to the load admittance. This assertion also holds for admittances.

Given a non-unity characteristic impedance, the transformation equation takes the form

$$\dot{z}\,(\lambda_g/4)\,\dot{z}\,(0) = z_{tr}^2 \tag{1.21}$$

Hence, a quarter-wave transformer with a characteristic impedance

$$z_{tr} = (r_1 r_2)^{1/2}$$

may be used to match a generator of internal resistance r_1 to an arbitrary purely resistive load r_2. Perfect match can only be achieved at the design frequency

$$\omega_0 = 2\pi v_p/\lambda_g$$

where v_p is the phase velocity in the transformer, and also at multiple frequencies when the transformer length is a quarter of the wavelength in the line plus an integer number of half-wavelengths. In the vicinity of the design frequency, a quarter-wave transformer gives an approximate match which degrades with increasing departure from the design frequency.

(3) **A half-wave transformer.** This term refers to a transmission-line segment having an electrical length equal to half the wavelength of the signal being transmitted or received, $\beta l = \pi$. With this electrical length, it follows from Eq. (1.20) that

$$\dot{z}\,(\lambda_g/2) = \dot{z}\,(0)$$

and

$$\dot{y}\,(\lambda_g/2) = \dot{y}\,(0)$$

That is, at the design frequency a half-wave line section of any z_{tr} does not transform load impedances (or admittances); it only converts them into themselves. At other frequencies, however, the input impedance does change, the change being progressively greater with increasing departure from the design frequency. This behaviour can be utilized to, say, compensate the frequency response of the load impedance.

Some applications of a half-wave transformer have as their basis the fact that *the normalized voltages at its input and output are equal in magnitude and opposite in sign.* This phase reversal follows, for example, from Eq. (1.9), subject to Eq. (1.17) at $\beta l = \pi$. This phase reversal is utilized in the widely used U-bend balun (Fig. 1.19) used in order to feed precisely in anti-phase a balanced $2R$ load from a coaxial line. The load impedance of the coaxial line at point A is, owing to the effect of the half-wave transformer, a parallel

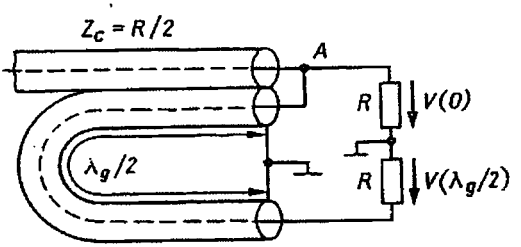


Fig. 1.19. U-bend type of balun

connection of two resistances of value R each. Because of this, it is exactly equal to the characteristic impedance of the cable, $Z_c = R/2$, and provides for a perfect match between the cable and load.

(4) **Measurement of load immitances by the Tatarinov method.** As has been observed by V.V. Tatarinov (USSR), the dimensionless impedance at any node of the longitudinal voltage distribution in a line is exactly equal to its TWR, that is, $\dot{z}_n = K$. Using this value as the load impedance in Eq. (1.18), we get

$$\dot{z}_{in} = 1/\dot{y}_{in} = (K + j \tan \beta l_n)/(1 + jK \tan \beta l_n) \qquad (1.22)$$

where l_n is the distance reckoned from any node in the voltage distribution on the line from the plane where $\dot{z}_{in}$ exists.

Equation (1.22) serves as the basis for an experimental procedure used to measure the input impedance of an unknown load with a so-called *slotted line.* This procedure is as follows. Moving the probe along the slotted line, the experimentor explores the longitudinal field distribution so as to locate v_{min} and v_{max} (see Fig 1.16). Next, he determines the travelling wave ratio*, $K = v_{min}/v_{max}$. Also, he measures the distance l_n from the node of the voltage distribution to the load input. Finally, using Eq. (1.22) or the impedance chart

* If a square-law detector is used in the slotted line, the TWR is found by extracting the square root of the ratio of the minimum to the maximum reading of the probe, $K = (\alpha_{min}/\alpha_{max})^{1/2}$. In precision work, it is essential first to determine the detector response.

(to be defined later), he finds the load impedance. When using Eq. (1.22), the distance l_n should be taken with a minus sign, because the load is located farther away from the generator than the voltage distribution node.

1.6. Transmission-Line Charts

In engineering practice, transformation of impedances and admittances in transmission lines (waveguides) can conveniently be handled by resort to any one of several charts or nomograms.

One of the simplest transmission-line charts is that due (in the Soviet Union) to A.R.Volpert*. It is based on Eq. (1.17) for the

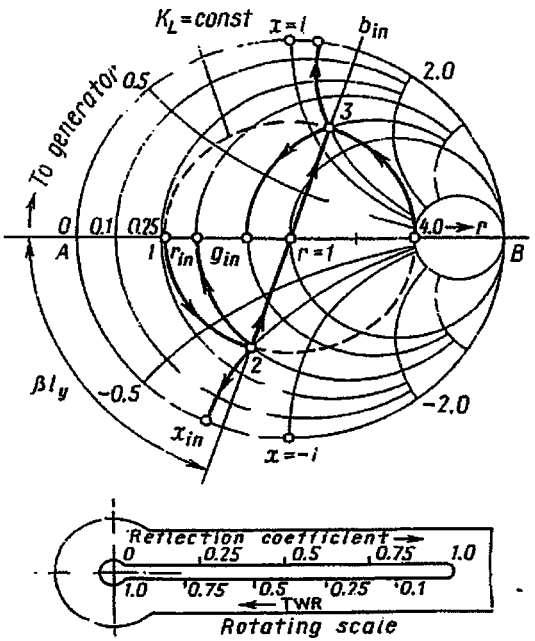


Fig. 1.20. Transmission-line chart

reflection coefficient and Eq. (1.18) for impedance transformation. On the transmission-line chart, the reflection is plotted in polar coordinates (Fig. 1.20), with the radial coordinate giving the magnitude of the reflection coefficient and the angular coordinate, φ, corresponds to twice the electrical length along the transmission line, that is, the phase angle of the reflection coefficient, $\varphi = 2\beta l$.

* Independently of Volpert and in the same year 1939, a similar chart was proposed by Phillip H. Smith of the United States. In fact, it is known as the Smith chart outside the Soviet Union.

For passive circuits, the magnitude of the reflection coefficient does not exceed unity, so the chart is bounded by a unity circle, $|\dot{\rho}| = 1$. At the centre of the chart, the reflection coefficient is zero. The electrical length is read from the chart as the azimuth of a radius vector rotating about the centre of the chart. Clockwise rotation represents the displacement of the selected node towards the generator. One complete revolution is equal to a half-wave distance along the line. A half-revolution is equal to a quarter-wave distance. It is usual to omit the lines of constant phase angles of the reflection coefficient, that is, radii, from the chart, and instead of lines of constant magnitudes of the reflection coefficient the chart shows the concentric circles of constant TWR (as dashed circles). As often as not, the transmission-line chart may be supplemented with a rotating radial scale which, in addition to TWR (or SWR) is graduated in units of the magnitude of the reflection coefficient. This scale is shown separately in Fig. 1.20.

Superimposed on the polar coordinate system for the reflection coefficient and TWR is a coordinate grid formed by the intersection of lines of constant normalized resistances r (or normalized conductance g) with lines of constant normalized reactance x (or susceptance b). The lines of constant r are circles with a radius equal to $1/(1 + r)$ and with centres located on the horizontal axis of symmetry. All circles of constant r come together at point B on the chart (see Fig. 1.20). Special mention should be made of the circle with $r = 1$, passing through the centre of the chart. The lines of constant x are likewise circles of radius $1/x$, with their centres located on the vertical line passing through point B. The lines of negative values of x lie in the lower half of the chart, and the lines of positive values of x in the upper half. On the horizontal axis of symmetry the reactances are equal to zero.

An example in the use of the transmission-line chart for measurements of load parameters by the Tatarinov method. Let the load TWR be equal to 0.25 and $l_n = 0.11\,\lambda_g$, that is, $\beta l_n \approx 40°$. The procedure for determining the load impedance $\dot{z}_{in}$ or the load admittance $\dot{y}_{in}$ is as follows.

(1) Referring to the impedance chart, locate on the horizontal axis running from short-circuit point A to the centre (this is the region of voltage nodes on the line) an impedance point corresponding to the node at the measured value of TWR (point *1*). At that point, $r_n = 0.25$.

(2) Find the circle of constant TWR equal to 0.25 and move around it from node point *1* to load point *2* (the heavy lines in Fig. 1.20), located at an electrical length $\beta l_n = 40°$. Remember to move counterclockwise (towards the load) because the load is farther away from the generator than the voltage node.

(3) Referring to the lines of constant r and x passing through load point *2*, read on the respective scales the real and imaginary parts of the impedance. In our case the result is $\dot{z}_{in} \approx 0.40 - j\,0.74$.

(4) To find the normalized admittance of the same load, recall that a quarter-wave matching element transforms $\dot{z}_{in}$ into $1/\dot{z}_{in}$ numerically equal to the admittance. Move from load impedance point *2* around the circle of constant $K_L = 0.25$ through a quarter-wave distance to arrive at a diametrically opposite point *3*, whose coordinates give the real and imaginary parts of the load admittance: $\dot{y}_{in} = 0.56 + j\,1.04$.

(5) Alternatively, the load admittance can be found by directly referring to an admittance chart. The only difference of this chart from an impedance chart is that the voltage nodes on the line are plotted along the right-hand half of the horizontal axis running from the centre to short-circuit point B (where g tends to infinity). The voltage node on the line where $K_L = 0.25$ corresponds to a node admittance, $\dot{y}_n = 1/K_L = 4.0$, lying naturally, on the circle of a constant TWR of 0.25. Moving round this circle from the nodal point through an electrical distance equal to $\beta l_n = 40°$ towards the generator leads us to the same point *3* which defines the load admittance.

1.7. Narrowband Matching of Transmission Lines

In Sec. 1.4 it has been shown that an optimum condition for a transmission line is a perfect impedance match between generator and load ($K = 1$). This condition can be achieved through the use of matching elements which are built solely from reactances (that is, they are free from ohmic losses) so as to preserve the high efficiency of the transmission system.

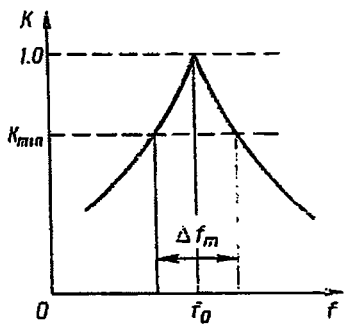


Fig. 1.21. TWR as a function of frequency in narrowband matching

Let us consider the so-called *narrowband matching* in which the travelling-wave condition ($K = 1$) exists only at a single design frequency. As the actual frequency departs from the design value, a mismatch occurs, and the TWR decreases. The typical frequency dependence of the TWR in the case of narrowband matching is shown in Fig. 1.21. The frequency band within which the TWR exceeds a specified minimum value, K_{min} (say, 0.8) is called the *match bandwidth*, Δf_m. In the case of narrowband matching, the

designer checks that this bandwidth is obtained by computation or experiment only after he has determined the values of all the matching elements and ignores it until that.

An alternative to narrowband matching is *broadband matching* in which case the values of the matching elements are chosen so as to obtain a maximum match bandwidth. In broadband matching, it is not required that the TWR should be unity at the design frequency, so it cannot serve as a basis for determining the values of the matching elements.

The relative match bandwidth, defined as the ratio of Δf_m to f_0 (where f_0 is the design frequency), is a function of the frequency response of the load and matching elements and the allowable mini-

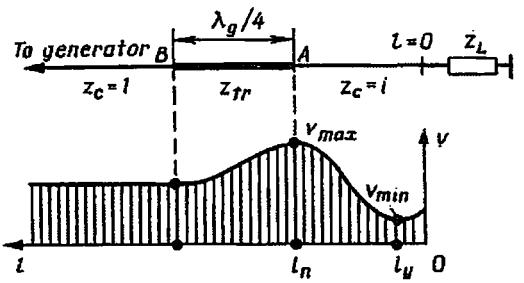


Fig. 1.22. Load impedance matching by a quarter-wave transformer

mum value of the TWR, K_{min}. This ratio may range between broad limits (from a few hundredths of one percent to several octaves) and has nothing 'narrow' about it.

For narrowband matching, it will suffice to balance out reflections from the load at the selected frequency by introducing an opposing reflection at some transverse plane on the line. For the match bandwidth to be as broad as practicable, the length of the matching element should be kept to a minimum and it must be located as close to the load as possible. If these requirements are satisfied, any departure from the design frequency will bring on a minimal change in the electrical length of the matching element, and the mismatch will increase at a slower rate with changes in frequency.

The devices most commonly used for narrowband matching are transmission-line segments with a variable characteristic impedance and also lumped reactances (say, stubs and diaphragms) located at the desired transverse plane.

Example 1.4. Matching by a quarter-wave transformer. The use of a quarter-wave machining element is illustrated in Fig. 1.22. The quarter-wave transformer of characteristic impedance $z_{tr} \neq 1$ takes up segment AB of the line. At the transformer output (point B), the impedance at the design frequency must be exactly

equal to unity (this is the match condition). Then, according to Eq. (1.21), the normalized impedance at the transformer input, point A, must likewise be purely resistive and equal to

$$r_A = z_{tr}^2$$

Therefore point A must be located either at a node or at a loop (an antinode) in the longitudinal voltage distribution. The latter case is shown in Fig. 1.22.

The resistance at a voltage node on the line is equal to the load TWR, K_L. On setting $r_A = K_L$, we obtain the required characteristic impedance of the transformer

$$z_{tr} = (K_L)^{1/2}$$

In the alternative case of point A being located at a loop of the voltage distribution (where the impedance is purely resistive and equal to $1/K_L$), the required characteristic impedance of the transformer will be

$$z_{tr} = 1/(K_L)^{1/2}$$

Thus, locating the input of a matching transformer at a loop calls for an increased characteristic impedance, whereas locating it at a node calls for a reduced characteristic impedance of the matching transformer. The choice depends on the value of the characteristic impedance that can be obtained by deforming the cross-sectional geometry of the transmission line.

In matching by a quarter-wave transformer, the TWR of the line varies as follows. It remains equal to K_L from load to the point of connection of the transformer, increases and becomes equal to $K_L^{1/2}$ within the transformer, and becomes equal to unity, $K_L = 1$, from the transformer output to the generator. A limitation of quarter-wave transformers is the difficulty of alignment and tuning after manufacture.

Example 1.5. Matching by lumped reactances. This form of matching (introduced in the USSR by Tatarinov in 1931) is widely used in all frequency bands from MF to EHF. Equivalent circuit diagrams for a reactance in series and in parallel with a transmission line are shown in Fig. 1.23. In each case, the travelling-wave condition is obtained through the choice of the connection point and of the nominal value of the reactance. This can best be done by reference to the transmission-line (impedance) chart.

A series matching reactance, jx_m, should be placed, as shown in Fig. 1.23*a*, at a distance l_x such that the real part of the impedance is exactly equal to unity. The imaginary part of the impedance at that point may be chosen such that it adds up to zero with x_m, thereby ensuring a perfect match at the design frequency. With these considerations in mind, enter the transmission-line chart with the load

impedance $\dot{z}_L$ (let it be point *1* in Fig. 1.24) and move round the respective circle of constant TWR (the heavy line in Fig. 1.24) until it cuts the line of constant resistance $r = 1$, thereby locating point

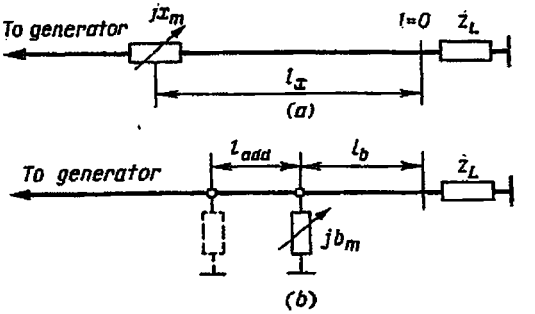


Fig. 1.23. Load impedance matching by (*a*) lumped series and (*b*) lumped shunt reactance

2. The angular distance βl_x between points *1* and *2* in Fig. 1.24 locates the connection point for a series reactance.

The reactance at point *2*, taken with a minus sign, gives the value of the matching reactive element to use, that is, $x_m = -x_2$. On ad-

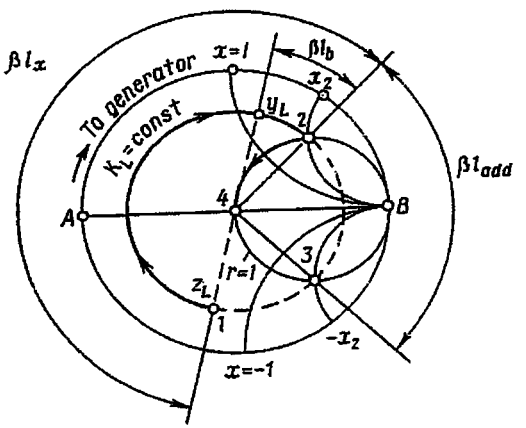


Fig. 1.24. To calculation of movable lumped-reactance matching elements

ding this reactance, the impedance at point *2* becomes equal to unity, which corresponds to moving round the $r = 1$ circle on the transmission-line chart to point *4*, or the point of perfect match.

Apart from point *2*, there is one more point where the circle of constant load TWR cuts the circle of unity resistance; this is point *3*. However, point *3* lies farther away from the load by an additional

distance, βl_{add}, so placing the matching reactance at point *3* might narrow the match bandwidth. On the other hand, this requires a reactance of the opposite sign — a factor which may turn out more attractive from structural considerations.

A *shunt reactive matching element* (Fig. 1.23*b*) may be examined in a similar way. However, instead of an impedance chart, one should use an admittance chart.

It follows from the foregoing that in the case of narrowband matching devices two controlled degrees of freedom (the location and characteristic impedance of a quarter-wave transformer, the location and nominal value of a reactance, etc.) will suffice in principle. Different arrangements of narrowband matching devices may be used for one and the same load. Which arrangement to use depends on the match bandwidth that can thus be obtained and structural limitations on the values of matching elements.

Chapter Two

Microwave System Components

2.1. Matched Loads

Matched loads or terminations, among the most commonly used microwave system components, serve to absorb the microwave energy transmitted on lines. Also, they can be used as dummy antennas in the alignment of transmitters and as standards of impedance in microwave measurements.

Matched loads are customarily described in terms of the magnitude of the reflection coefficient, $|\dot{\rho}|$, or the corresponding value of TWR or SWR within the specified bandwidth. It is technically feasible to build matched loads with $|\dot{\rho}| \leqslant 0.01$ for a relative bandwidth $\Delta f_m/f_0$ of 20 or 30% or more. Because the magnitude of the reflection coefficient is small, no special requirements are imposed on the phase of the load reflection coefficient, and this phase may range anywhere between zero and 2π.

Another important consideration as regards matched loads is the power dissipation rating. Some matched loads are built to dissipate low levels of power (not over 1 W), while others can handle large blocks of power.

The physical design of a matched load depends on the type of line (waveguide) in which it is to be used and the power level expected. Also, matched loads may be lumped and distributed, and the latter may be designed for very high levels of power by increasing their size and mass.

For *coaxial lines*, the simplest matched load is a lumped resistor with a resistance equal to the characteristic impedance of the line. In the SHF band, however, such a resistor becomes comparable in size with the guide wavelength, the input resistance becomes frequency-dependent, and the quality of match is impaired appreciably. As a way of bringing down the reflection coefficient and extending the operating bandwidth, coaxial matched loads for the SHF band are often fabricated in lossy non-uniform transmission-line sections. The absorbing element in such a load may be a block of suitable

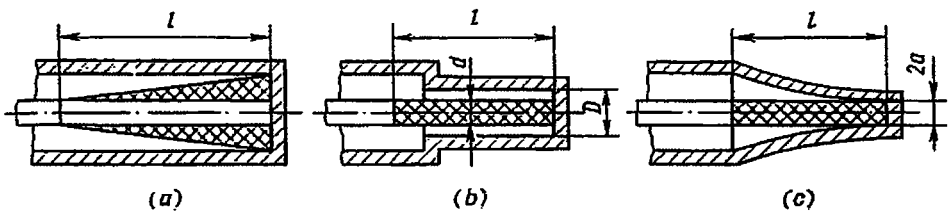


Fig. 2.1. Coaxial matched loads (terminations)

material or a thin film. A matched load with an absorbing cone-shaped block is shown in Fig. 2.1*a*. A good quality of match with this type of load is obtained when the length of the absorbing element is equal to or greater than the wavelength on the line, $l \geqslant \lambda$.

More often, coaxial matched loads use an absorbing element in the form of a ceramic cylinder coated by a metal oxide or conducting carbon film. Film thickness is made small in comparison with the skin depth, so the surface resistance of the film is almost independent of frequency. To make the input impedance of a matched load with a cylindrical absorbing element purely resistive and nearly constant over a broad bandwidth, it is fitted with a nonuniform metal shield suitably shaped and proportioned.

Figure 2.1*b* shows a coaxial matched load with a step-shaped shield. It has been found that an optimal quality of match at $\lambda \geqslant 6l$ is obtained when the shield diameter is reduced so that

$$138 \log (D/d) = Z_c / \sqrt{5}$$

where Z_c is the characteristic impedance of the main coaxial waveguide. The length of the step on the outer conductor must be a little shorter than that of the absorbing film.

In broadband coaxial loads, the outer shield is given the shape of a funnel (Fig. 2.1*c*).

Matched loads for strip and microstrip lines are thin-film strips of a resistive material applied to a circuit board and shorted at one end to the ground plane of the line. The film thickness is chosen to be a fraction of the skin depth, and the film length may be small in com-

parison with the wavelength on the line. Because of poor heat transfer, however, such lumped loads can stand up to low power levels only. For a greater power dissipation, it is usual to fabricate loads in long ($l \approx \lambda$) uniform or nonuniform lossy line sections. This, however, requires that the absorbing surface be given a suitable shape. Microwave strip-line circuits also use discrete matched loads in the form of ceramic strips or posts coated with an absorbing material. When matched loads are fabricated on strip-line boards, it is often

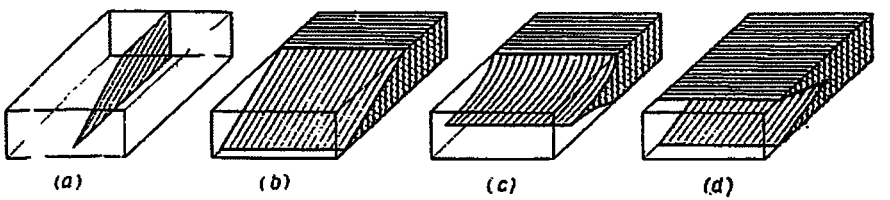


Fig. 2.2. Waveguide matched loads (terminations)

difficult to short-circuit conductors to the ground planes. With a relative match bandwidth of $\Delta f_m/f_0 = 5$ to 8% this difficulty is circumvented by using quarter-wave stubs which have a nearly zero input impedance.

Waveguide matched loads are mostly made in shorted waveguide sections enclosing a strip or a block of some absorbing material varying in profile. In low-power matched loads, the absorbing element is in the form of a dielectric plate coated by a graphite or metallic film (Fig. 2.2*a*). Plug-type loads (Fig. 2.2*b* through *d*) intended for high power levels are made of composite materials based on powdered graphite, carbonyl iron, or silicon carbide. The plugs may be shaped as a wedge or a pyramid to reduce reflections. If the attenuation due to an absorbing element is in excess of 20 or 25 dB, reflections from the final short will be negligibly small. For better heat withdrawal, the area over which the absorbing element comes in contact with the waveguide walls is made as large as possible, and the outer surface of the waveguide is fitted with a radiator.

2.2. Reactive Loads

Reactive loads or terminations used as standards in microwave measurements and in matching and control microwave devices are required to have a stable normalized input impedance which can accurately be computed from physical dimensions. Most often, reactive loads are shorted segments of closed-wire transmission line, commonly known as *shorted stubs*. The key parameter of a real stub is the input SWR which should be as high as possible. The SWR of nonadjustable coaxial or waveguide stubs with a fixed shorting plunger

soldered in place may be as high as 500 or even more. The SWR of adjustable stubs with a movable shorting plunger is lower because of additional losses at contacts, but is, as a rule, in excess of 100. The no-load condition, that is, an open-circuited output in stubs can only be realized in shielded multi-wire lines where no spurious emission of radiation can take place.

Several designs of movable shorting plungers for rectangular waveguides are shown in Fig. 2.3, with the sectional views arranged to

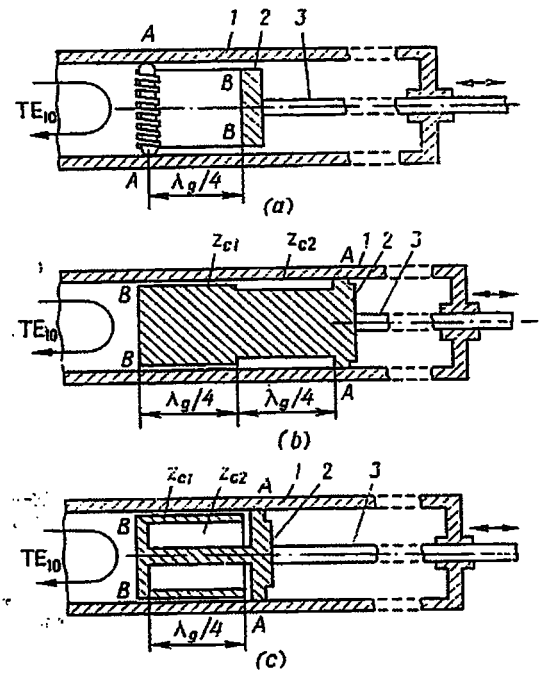


Fig. 2.3. Waveguide shorting plungers: *1*, waveguide; *2*, plunger; *3*, connecting-rod

be parallel to the narrow dimension of the waveguide. In the design of Fig. 2.3*a* the split contact springs *A* are moved away from the shorting wall *B* for a distance of $\lambda_g/4$ inside the waveguide. Because of this, the contacts are located at a waveguide section with zero longitudinal current on the waveguide walls, and any departure of the contacts from an ideal condition cannot lead to an increase in power dissipation.

In the design of Fig. 2.3*b*, the contacts *A* are connected into the waveguide via two impedance-transforming line sections with low values of the normalized characteristic impedance, z_{c1} and z_{c2}.

Assuming that the contact resistance at point A is r_A and applying twice the equation for impedance transformation by a quarter-wave transformer, we obtain the input impedance at point B:

$$r_B = r_A (z_{c1}/z_{c2})^2$$

On choosing z_{c1} to be substantially smaller than z_{c2}, it is possible to markedly reduce the equivalent contact resistance r_B and to increase the SWR of the plunger.

In the design of Fig. 2.3*c*, the points of contact are located in the middle of a folded shorted half-wave line section made up of two cascaded quarter-wave sections with z_{c1} and z_{c2}. To the contact resistance r_A is added the infinite reactance of the shorted quarter-wave stub with a characteristic impedance z_{c2}, and the sum of the impedances of the contact and stub is transformed by the quarter-wave stub with a characteristic impedance z_{c1} into a practically zero resistance at point B. (In other words, a virtual short-circuit is established for microwave currents at point B.)

The designs we have just examined are quite applicable to coaxial plungers in the SHF band.

2.3. Supports for Coaxial Microwave Transmission Systems

Rigid coaxial microwave transmission systems pose the problem of supporting the inner conductor. These supports have to meet a number of conflicting requirements: they ought not to give rise to re-

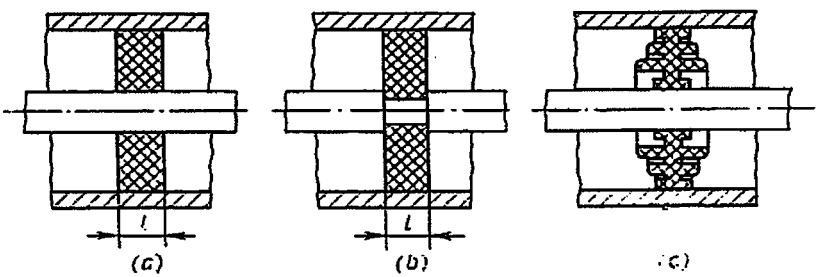


Fig. 2.4. Dielectric disc supports for a coaxial line

flections, impair electrical strength, increase the attenuation, or reduce the operating bandwidth. Also, they must permit repeated disassembly and assembly of the transmission system, etc.

The inner conductor is most often supported by dielectric discs or by metal insulators.

A simple *dielectric support* (Fig. 2.4*a*) connected in a coaxial guide may be represented by a line section of an equivalent length

$$l_{eq} = l \sqrt{\varepsilon_r}$$

and with a reduced characteristic impedance. At low frequencies ($l_{eq}/\lambda \ll 1$), simple dielectric supports have an insignificant effect on the line performance. At higher frequencies, however, the electrical length of the support increases, and so does the attenuation. Reflections can be minimized by reducing the diameter of the inner

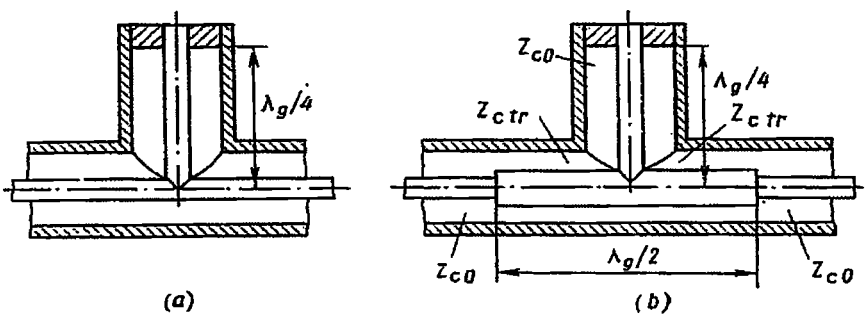


Fig. 2.5. Metal supports for a coaxial line

conductor within the length taken up by a disc, thus keeping the characteristic impedance unchanged. This leads to the design shown in Fig. 2.4*b* where what is called the *flush* or *sunk disk* may be split across the diameter for ease of assembly.

Dielectric supports reduce the electric strength of the waveguide or line because it is now easier for a breakdown to occur at the surface of the support and also because the field intensity in the small, but inevitable gap between the support and the inner conductor is very high. The electric strength can be enhanced by making concentric grooves and ribs in the surface of the supports as shown in Fig. 2.4*c*.

In the SHF band, dielectric supports give way to *metal insulators*—rigid, parallel shorted quarter-wave stubs which support the inner conductor as shown in Fig. 2.5*a*. A metal support does not disturb the match at the operating wavelength, reduces the electrical strength only very slightly, and introduces but a very small excess loss. Unfortunately, the simple metal support of Fig. 2.5*a* has a limited bandwidth because variations in frequency lead to changes in the electrical length of the support and introduce a mismatch in the system.

What is known as a *broadband metal support* (Fig. 2.5*b*) has, in addition to a quarter-wave metal insulator, a half-wave transformer which has a reduced characteristic impedance, $Z_{c,\ tr}$. The value of $Z_{c,\ tr}$ may be chosen such that the support is ideally matched not only at the centre frequency, but also at two more frequencies, one above and the other below the centre frequency of the band. The

oscillating SWR-vs-frequency curve for a matched microwave transmission system with a broadband support for $Z_{c.tr} = 0.8$ is shown in Fig. 2.6*a*. As $Z_{c.\,tr}$ is varied, the humps of the SWR curve and the relative match bandwidth $\Delta f_m/f_0$ at this level change as shown in

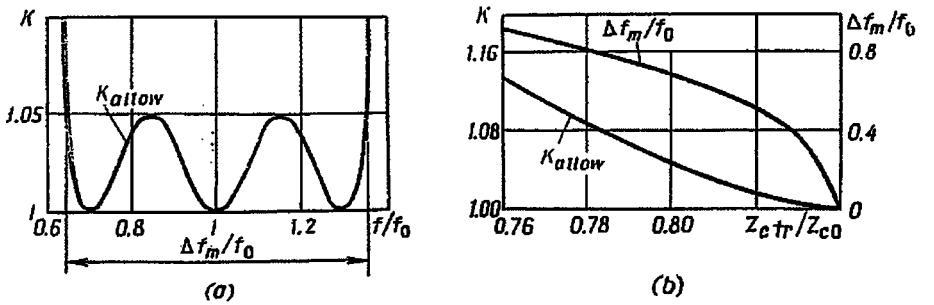


Fig. 2.6. SWR-vs-frequency response and parameters of a broadband metal support

the plot of Fig. 2.6*b*. The operating bandwidth of a broadband metal support at an allowable SWR level of 1.1 is 80% of the centre frequency.

2.4. Microwave Connectors and Couplings

For ease of assembly and disassembly, the components of a microwave transmission system are fitted with suitable connectors, couplings, or adapters. They are designed to provide a reliable electric contact at the joints, to preserve the quality of match and electrical strength of the system, to introduce a minimum excess signal attenuation, and not to give rise to spurious emission of radiation.

In microwave connectors for flexible coaxial cables, electric contact is provided by spring collets and jacks (Fig. 2.7*a*) held together by outer threaded collars, nuts or similar fasteners. At any section of a coaxial microwave connector the diameters of the conductors are chosen so as to maintain the characteristic impedance of the line at a constant value with allowance for the effect of the dielectric. The quality of match in microwave coaxial connectors is to a marked degree dependent on how well the cable is terminated in the connectors. With good workmanship, the rms SWR is of the order of 1.05 to 1.15.

Microwave couplings for rigid high-power coaxial waveguides have no dielectric supports (Fig. 2.7*b*). Rectangular waveguide segments are joined by flange couplings which may be of either the plane-faced or cover type and the choke type.

A *plane-faced* or *cover flange coupling* has flat contact surfaces finished off square and smooth and tightly clamped together. This type of coupling can give a good joint, but its quality rapidly deteriorates with each assembly and disassembly. For better contact, the mating surfaces have pins which carry a bronze gasket with several leaf springs fitting snuggly against the inside surfaces of the waveguide

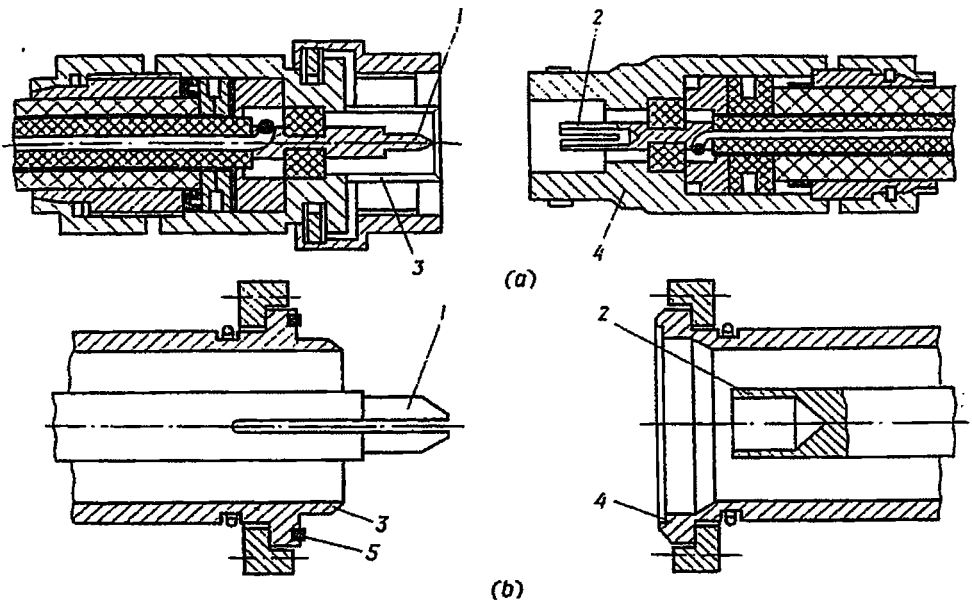


Fig. 2.7. Microwave coaxial connectors: *1*, male contact; *2*, female contact; *3*, plug; *4*, receptacle; *5*, seal

segments joined (Fig. 2.8). To keep out dust and moisture, the flanges have grooves on either side of the bronze gasket, which receive rubber seals.

In a *choke flange coupling*, shown in sketch form in Fig. 2.9, contact between the joined waveguide segments is provided by a series shorted half-wave stub made in the form of grooves and recesses inside the flange. The quarter-wave section between the short-circuit point A and the mechanical contact point B acts as a TE_{11} coaxial waveguide, while the other quarter-wave section between the mechanical contact point B and the point C where the stub is connected into the waveguide acts as a radial-line section. The point of mechanical contact occurs at a node of surface current J, so very little power is dissipated across the contact resistance r. A virtual short-circuit between the joined waveguide sections at the point C is established because the total length of the choke grooves from point A

to point C is half the guide wavelength. The insides of the waveguide are sealed off by a gasket dropped in an additional concentric groove. Choke flange couplings will do their job well even if the two pieces of guide are slightly separated or misaligned, and will not impair the electric strength of the system. Unfortunately, their SWR varies markedly with frequency and their design is rather complicated.

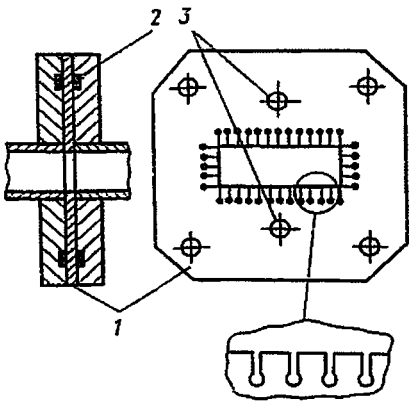


Fig. 2.8. Waveguide plane-faced flange coupling: *1*, contact gasket; *2*, grooves and packing; *3*, locating-pin holes

Rotary (rotating) joints, also known as *rotary couplers*, permit one section of a waveguide to rotate with respect to another without disturbing electric contact or the quality of match.

Rotary joints use short sections of a TEM coaxial waveguide or a TM_{01} circular waveguide. In circular waveguides, the rotary joint may also use the circularly polarized TE_{11} mode.

Some rotary joints use rubbing contacts, but they are unreliable where rotation is continuous and the transmitted power is maintained at a high level. A better choice is a rotary joint with a choke groove, but its performance varies with frequency.

In a *coaxial rotary joints* (Fig. 2.10*a*), choke grooves are provided in both the outer and the inner conductor. Each choke is a series, two-

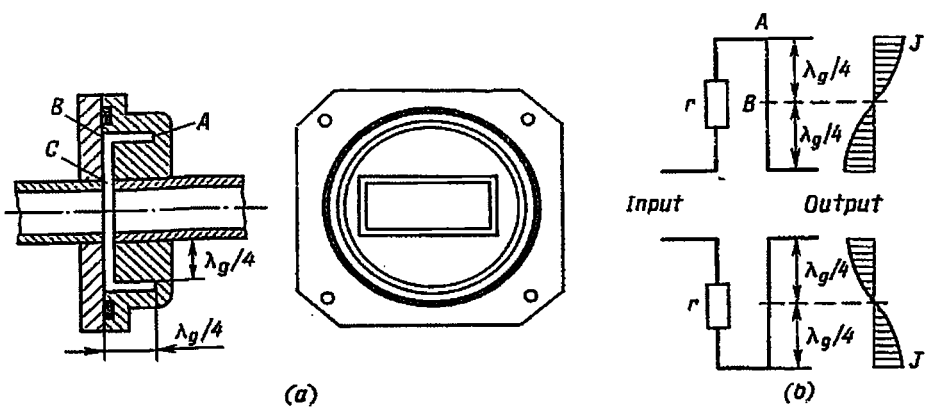


Fig. 2.9. Waveguide choke-flange coupling; *a*, sketch; *b*, equivalent circuit

step shorted half-wave stub. The equivalent circuits for the chokes in the outer and inner conductors are the same and shown separately

in Fig. 2.10*b*. (Note that similar equivalent circuits apply to the choke grooves in the waveguide plunger of Fig. 2.3*c* and in the waveguide flange coupling of Fig. 2.9*a*.) Inside the choke a standing electromagnetic wave is formed, with its current node located at the rubbing contacts which lie within a quarter of the operating wavelength, λ_0, of the shorted ends of the stubs. Therefore, the voltage

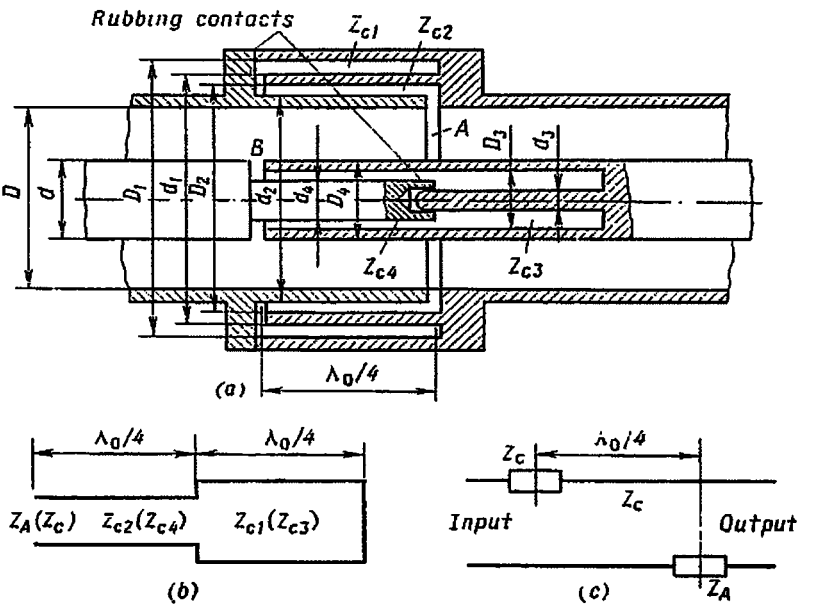


Fig. 2.10. Coaxial choke-type rotating joint: *a*, sketch; *b*, equivalent circuit of the choke groove; *c*, equivalent circuit of the joint

drop across the contact resistances is zero, no microwave power is lost, and the performance of the rotary joint at the operating frequency is independent of the quality of the rubbing contacts.

When the frequency is other than the design operating value, the input impedances of the choke, Z_A and Z_B, are non-zero, and the rotary joint gives rise to reflections in the transmission system. For small amounts, Δf, away from the design operating frequency, f_0, the input impedance of each choke can be found by treating it as a transmission-line section with an electrical length equal to

$$\beta l = (1 + \Delta f/f_0)\pi/2$$

which is loaded into another short-circuited line section with an input impedance equal to

$$Z_L = jZ_{c1} \tan \beta l$$

On setting

$$\tan \beta l = \tan [(1 + \Delta f/f_0)\, \pi/2] \approx 2f_0/\pi\Delta f$$

we obtain after simple manipulations

$$X_A \approx (Z_{c2}/Z_{c1})\, (Z_{c2} + Z_{c1})\, (\pi/2)\, (\Delta f/f_0)$$

$$X_B \approx (Z_{c4}/Z_{c3})\, (Z_{c4} + Z_{c3})\, (\pi/2)\, (\Delta f/f_0)$$

It follows from the expressions for X_A and X_B that reflections at a rotary joint can be reduced by bringing down the characteristic impedances Z_{c2} and Z_{c4} of the open-circuited choke grooves and by increasing the characteristic impedances Z_{c1} and Z_{c3} of the short-circuited choke grooves.

Referring to the equivalent circuit diagram of a rotary joint in Fig. 2.10*c*, the input TWR of the joint in the case of an ideal match at the output is found to be

$$K \approx 1 - (\pi/2Z_{c,m})\, (\Delta f/f_0)\, [Z_{c2}\, (1 + Z_{c2}/Z_{c1}) - Z_{c4}\, (1 + Z_{c4}/Z_{c3})]$$

where $Z_{c,m}$ is the characteristic impedance of the main transmission line and Δf is the amount of detuning from f_0. The second term in the square brackets is subtracted from the first because the gaps in the inner and outer conductors are spaced a quarter of the guide wavelength apart and because the respective reflections cancel one another. For characteristic impedances that practical choke grooves may have, the TWR of a rotary joint, as follows from the equation for the TWR, exceeds 0.9 over a bandwidth which is 50 to 70% of the operating frequency.

The electrical strength of a coaxial choke-type rotary joint is somewhat lower than that of the line because the maximum electric field strength in the middle of the choke groove of the inner conductor is greater than the maximum electric field strength at the surface of the central conductor of the main waveguide.

Choke-type rotary joints fabricated in circular waveguide can be analysed in a similar way. However, there is only one choke to reckon with, and so Z_{c4} in the equation for the TWR may be set equal to zero.

2.5. Transitions. Mode Transducers

In microwave systems it is often necessary to change from one form of waveguide to another or from a waveguide to a coaxial line or back. This job is done by *transitions*, also known as *adapters*. Where it is desired to change from one mode in one guide to another mode in another, *mode transducers* are used.

When designing transitions and mode transducers, the primary objective is to obtain a high quality of match over the specified bandwidth and ample electrical strength.

Consider several forms of transitions.

A TE_{10} mode is launched in a rectangular waveguide from a TEM coaxial line by a *coaxial-to-waveguide transition*, *adapter*, or *transducer*

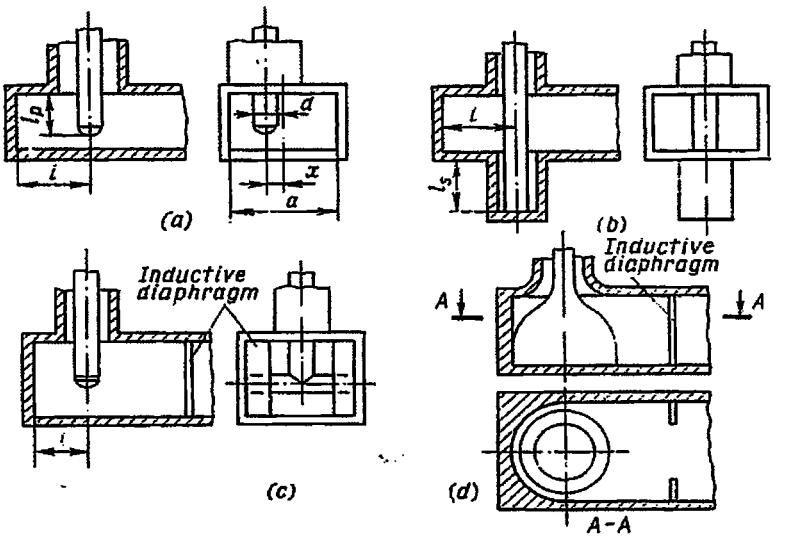


Fig. 2.11. Coaxial-to-waveguide transitions: *a*, probe type; *b*, coaxial-stub type; *c*, transverse-post type; *d*, doorknob type

(Fig. 2.11). The key element in such a device is a probe located parallel to the electric lines of force, E, in the waveguide shorted on one side.

In the *probe type of transition* (Fig. 2.11*a*), the inputs are matched by adjusting the probe length l_p, and also the distances l and x which define the location of the probe. For a greater match bandwidth, it is advisable to increase the probe diameter d. When well made, a probe transition gives a match bandwidth of 15 to 20% of the design frequency at a TWR of at least 0.95. A limitation of probe transitions is a reduction in electric strength because of the concentration of electric lines of force at the probe tip. To some extent this shortcoming is avoided in a *series-stub coaxial-to-waveguide adapter* (Fig. 2.11*b*), but even with the most careful choice of l and l_s, the effective bandwidth here is a mere 7% of the design (centre) frequency.

Better performance in terms of match and electric strength is shown by an *adapter with a transverse post* (Fig. 2.11*c*) supplemented

with an inductive matching diaphragm (or iris). In this design, the relative match bandwidth is about 15%. A maximum bandwidth (around 20% at a TWR of at least 0.95) and high electrical strength are obtained in the *doorknob type of coax-to-waveguide transition* (Fig. 2.11*d*). They, however, call for a careful choice of conductor shape (and this can only be done by trial and error), and a further adjustment for the quality of match with an inductive diaphragm.

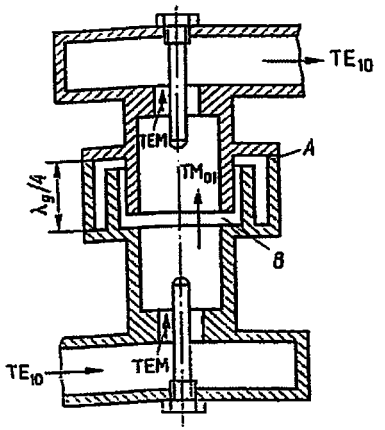


Fig. 2.12. TM_{01} waveguide rotating joint

The manner in which a TM_{01} mode is launched by a coax-to-waveguide transition in a circular waveguide is illustrated in Fig. 2.12, taking a rotary joint as an example. Short segments of a coaxial TEM waveguide filter out the higher modes and prevent the launching of a spurious TE_{11} mode in the circular waveguide. (The TE_{11} mode is of a lower order than the TM_{01} mode.) The rotating sections of the circular waveguide are joined by a half-wave coax choke similar in design to the chokes used in the coaxial rotary joint of Fig. 2.10.

The *lower-order* TE_{11} *mode* can be launched in a *circular waveguide* by a tapered transition whose cross-section gradually changes from rectangular to circular (Fig. 2.13*a*). If such a transition has a taper

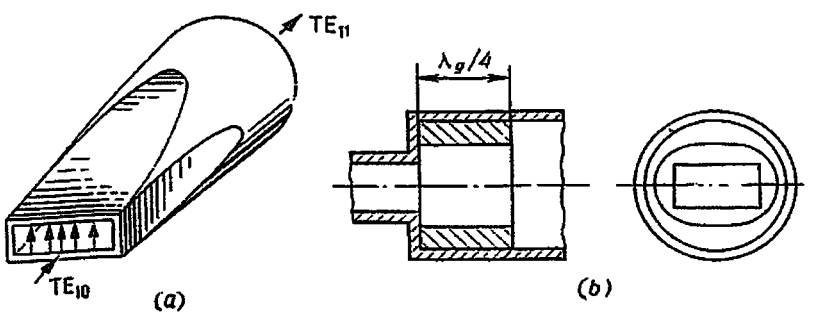


Fig. 2.13. Tapered and in-line transitions from TE_{10} rectangular guide to TE_{11} circular guide

longer than the guide wavelength, reflections will be insignificant over a broad bandwidth. A more compact narrow-band transition is shown in Fig. 2.13*b* where a rectangular and a circular waveguide are joined via an oval quarter-wave matching transformer.

In a circular waveguide, the TE_{11} mode can alternatively be launched from a rectangular waveguide through a hole in the side wall. If the broad faces of a rectangular waveguide are arranged to run parallel with the axis of a circular waveguide (Fig. 2.14*a*), the circular waveguide will be excited in the TE_{11} mode which will propagate in the same phase in both directions from the junction. If the coupling slot is made across a circular waveguide (Fig. 2.14*b*), the TE_{11} modes

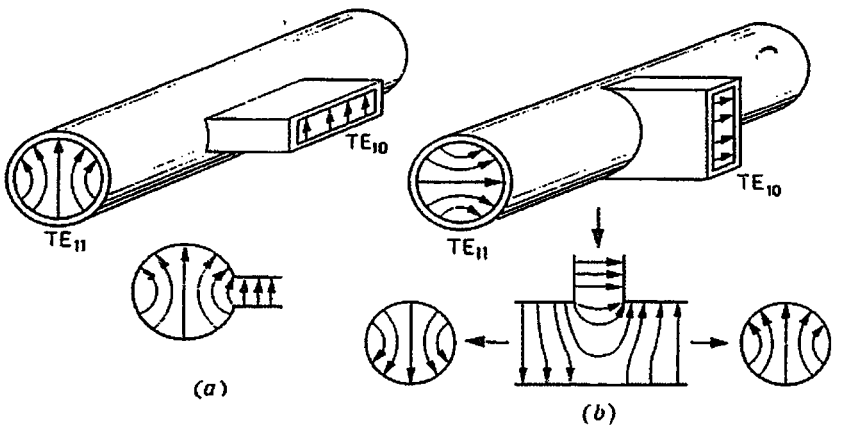


Fig. 2.14. Tee-junctions of rectangular and circular waveguides

excited on the right and left of the slot will be in antiphase. If it is required to transmit a TE_{11} mode in one direction only, one end of the circular waveguide must be shorted. For the junction shown in Fig. 2.14*a* the distance between the middle of the slot and the short must be close to a quarter of the guide wavelength, while in the junction shown in Fig. 2.14*b* it must be close to a half of the guide wavelength.

Consider the manner in which it would be possible to launch an *axially symmetrical* TM_{01} *mode in a circular guide* from a TE_{10} rectangular guide without resort to coax-to-waveguide adapters. In the arrangement of Fig. 2.15*a*, a rectangular guide is joined to a coaxial one by way of a transverse slot. For better excitation of the TM_{01} mode the circular guide is shorted on one side within a half of the TM_{01} guide wavelength from the coupling slot. In order to suppress the spurious TE_{11} mode which is likewise excited by the coupling slot, the shorted section of the circular guide is fitted with a slim metal ring. The perimeter of the ring is chosen to be close to λ_0 so that the TE_{11} mode could give rise to resonant waves with a single variation round the perimeter. This ring acts on the TE_{11} mode as a short. Located within a quarter of the TE_{11} guide wavelength of the slot centre, the ring can effectively suppress the TE_{11}

mode in the circular guide. The resonant ring has practically no effect on the TM_{01} mode in which the electric lines of force are radial and lie in the plane of the resonant ring.

Another form of a TM_{01} mode launcher in a circular guide with a high degree of suppression of the spurious TE_{11} mode is shown in Fig. 2.15*b*. The rectangular guide is joined to the circular one in the same manner as in the previous design, with one half of the circular guide shorted right at the junction. Also, the circular guide

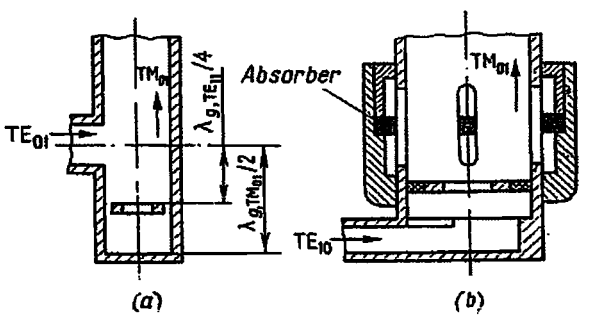


Fig. 2.15. Methods of launching a TM_{01} mode in a circular guide

encloses a resonant ring which shorts the guide for the TE_{11} mode. This mode, leaking through the resonant ring, is absorbed and excites the TE_{011} coaxial resonator through longitudinal slots in the walls of the circular guide. At a loop of the electric field in the resonator, there is an absorbing ring in which the energy of the TE_{11} mode is dissipated. The TM_{01} mode does not produce transverse currents in the walls of the circular guide and cannot therefore excite the longitudinal slots and the absorbing resonator.

It is especially difficult to design a TE_{01} *mode launcher in circular guide*. Here, the primary objective is to maintain high purity for the TE_{01} mode and to suppress as much as practicable a whole range of the higher and lower modes capable of propagating in a larger-diameter circular guide. A likely design for a transition from a TE_{10} rectangular guide to a TE_{01} circular guide is shown in Fig. 2.16. As is seen, it is basically a tapered rectangular-to-circular transition. An E-plane tee and two 90-degree longitudinal twists oriented in opposite directions make up a system of two rectangular guides which are coupled on their narrow sides and support fields equal in amplitude but opposite in phase. This system gradually changes into two sectorial guides having a common web. As the aperture of the sectorial guide increases, a circular guide with a longitudinal metal partition is formed. A break in this partition does not change the field pattern, and the output section is a TE_{01} circular guide. For high pu-

rity of the TE_{01} mode, this form of mode transducer must be of a length $l \gg \lambda_0$.

Some difficulties related to the quality of match over a broad bandwidth also arise in making transitions from strip to coaxial and rectangular waveguides.

Coaxial-to-strip line transducers. These fall into two broad classes. In one, both lines are arranged so that their axes are on the same li-

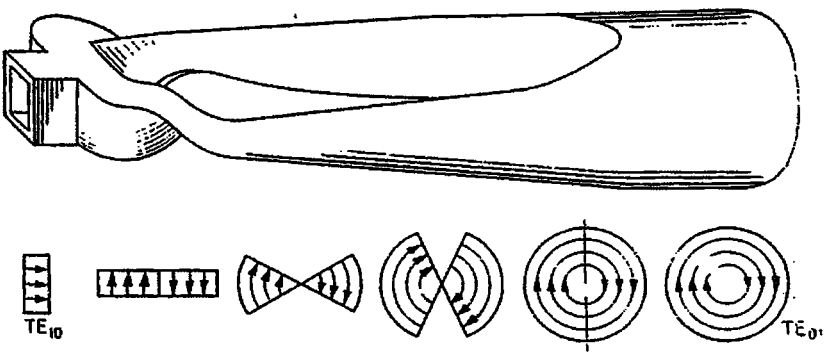


Fig. 2.16. Tapered transition used to launch a TE_{01} mode in a circular guide

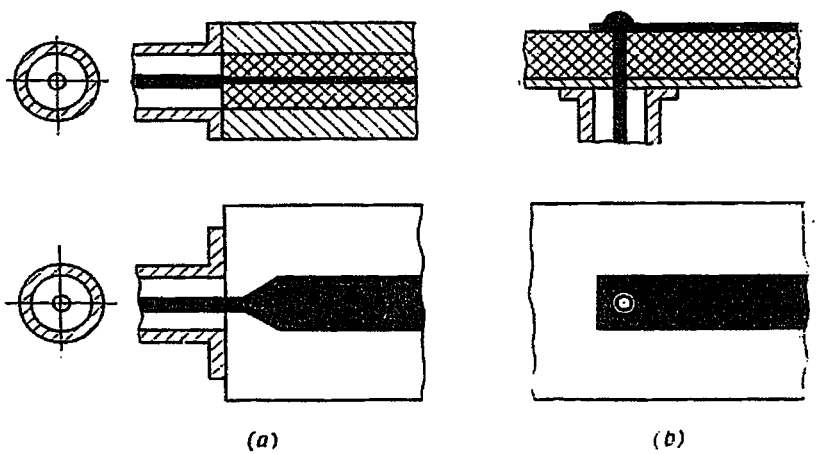


Fig. 2.17. Coaxial-to-microstrip transducers

ne, that is, coaxial or in-line. In the other, one line is at right angles to the other (Fig. 2.17). To reduce discontinuity at the joint the inside diameter of the outer conductor in the coaxial line must be close in value to the spacing between the outer ground planes of a shielded (or balanced) strip line or to twice the thickness of the base

ground plane of a microstrip line. The desired quality of match in the case of an in-line transducer is ensured by giving a taper to the strip (or microstrip) line as shown in Fig. 2.17*a*. Where the coaxial line is at right angles to the strip (or microstrip) line (Fig. 2.17*b*), the required quality of match is obtained by adjusting the diameter of the coupling post extending through the dielectric substrate and

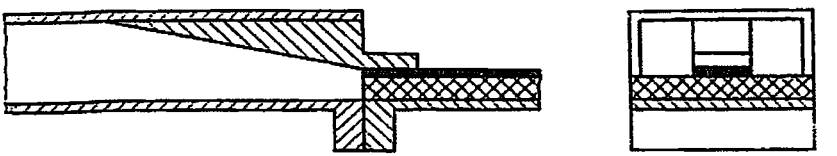


Fig. 2.18. Waveguide-to-microstrip transducer

also by adjusting the dimensions of the coaxial diaphragm at the output from the coaxial guide and of the open-circuited stub fabricated in strip (microstrip) line section. As often as not, coaxial-to-strip line transducers are combined with coaxial connectors.

Waveguide-to-strip line transducers. These devices are intended to excite a strip (or a microstrip) line from a TE_{10} rectangular guide. A strip (or microstrip) line can be joined to a rectangular guide by a tapered or stepped transformer fabricated in single-ridged guide (Fig. 2.18). This design provides broadband matching between rectangular guides and strip (microstrip) lines and prevents spurious emission from the open end of the guide.

2.6. Waveguide Corners, Bends and Twists

Waveguide corners, bends and twists fall in a class of discontinuities which impair the quality of match and electric strength.

Whatever the form of the microwave guide system, corners tend to launch non-propagating higher modes whose fields can store some electromagnetic energy. To minimize the resulting reflections, corners are often supplemented with various matching devices. For example, a 90-deg bend fabricated in coaxial line can be combined with a quarter-wave support and supplemented with a shallow groove milled on the inner conductor (Fig. 2.19*a*). With a judicious choice of position and dimensions for the groove and of length for the quarter-wave support, a good quality of match can be maintained over a broad bandwidth.

For corners the quality of match can be enhanced by giving them a miter cut on the outside as shown in Fig. 2.19*b* for a length equal to one-third of the distance along the diagonal between the inner and outer apexes of a sharp corner. However, even such modified cor-

ners will introduce some excess delay which must be allowed for in calculating the electrical lengths of resonant sections.

Miter cuts are effective in reducing reflections in rectangular and circular guides as well. Optimal dimensions for miter cuts (Fig. 2.19*c*) can be found from plots in relevant reference sources.

Corners are abrupt by nature and tend to concentrate electric lines of force, which fact impairs the electric strength of the transmis-

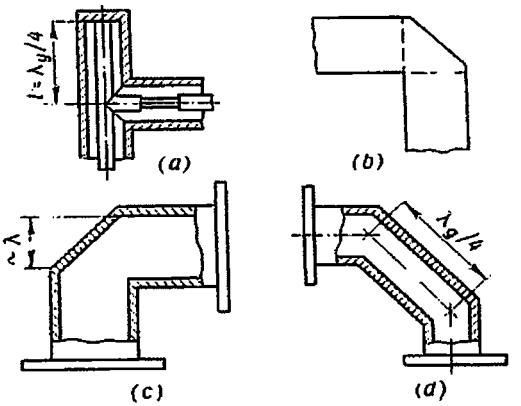


Fig. 2.19. Miter-cut and two-cut bends for waveguides

sion system. This drawback is to a considerable extent minimized in two-cut miter elbows and bends. In a two-cut miter elbow (Fig. 2.19*d*), the two discontinuities (cuts) are spaced apart a distance equal to a quarter of the guide wavelength. The quality of match is improved because reflections from each cut are reduced and also because reflections from the two cuts tend to cancel one another.

Bends can be represented by the equivalent circuit diagram for a line section with a somewhat changed characteristic impedance. For a better quality of match, long-radius bends should preferably be used or the length of a bend should be made a multiple of a half of the guide wavelength.

2.7. Reflecting Discontinuities in Waveguides

Apart from spurious discontinuities inevitable in any practical microwave transmission system (joints, bends, corners, twists, transitions, and the like), which are all detrimental to its performance, it is common practice to introduce deliberate discontinuities as 'building blocks' for matching devices, resonators, filters, and other microwave components. Some of them are discussed below.

Butt joints of dissimilar guides. In stepped transitions and filters, the two component rectangular waveguides differing in cross-sectional area are often joined end-to-end and in line, as shown in Fig.

2.20. Such a butt joint can be represented by the equivalent circuit diagram of Fig. 2.20*b* with normalized characteristic impedances z_{c1} and z_{c2}. Their ratio is given by

$$z_{c1}/z_{c2} \approx (a_2/a_1)\,(b_1/b_2)\,(\lambda_{g1}/\lambda_{g2}) \qquad (2.1)$$

where a and b are the width and height of guide, and λ_g is the guide wavelength. The numerical subscripts identify which guide is meant.

A more exact equivalent circuit diagram includes the shunt capacitance C due to the existence of non-propagating modes near the

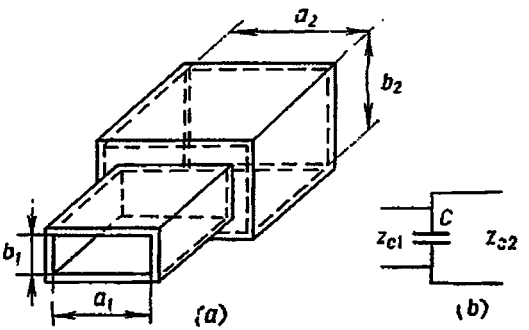


Fig. 2.20. Butt joint of two dissimilar rectangular guides: *a*, sketch; *b*, equivalent circuit

joint. Its value can be looked up in reference sources. The effect of this shunt capacitance on the performance of a stepped transformer can be balanced out by reducing the length of the component sections.

Waveguide diaphragms. A *waveguide diaphragm* (also called an *aperture*, *window*, or *iris*) consists of two thin metal plates that partly block the cross-section of the guide. Rectangular waveguides mostly use symmetrical inductive, symmetrical capacitive, and resonant diaphragms (Fig. 2.21).

In an *inductive diaphragm* (Fig. 2.21*a*), the transverse currents on the broad faces of the guide have some of their path completed through the diaphragm plates joining the faces. The magnetic fields due to these currents store some magnetic energy. Such a diaphragm can be represented by an equivalent circuit made up of an inductive susceptance which shunts the guide at the location of the diaphragm. Approximately, the normalized susceptance of an inductive diaphragm, b_L, is given by

$$b_L \approx -\,(\lambda_g/a)\cot^2(\pi d_L/2a)$$

where d_L = width of the diaphragm opening
a = broad dimension of the guide
λ_g = guide wavelength

A *capacitive diaphragm* (Fig. 2.21*b*) reduces the opening between the broad faces of the guide. The electric field E is concentrated between the diaphragm edges, and some electric energy is stored there. Therefore, a capacitive diaphragm has an equivalent circuit consisting of a capacitive susceptance shunting the guide at the location of the diaphragm. The normalized capacitive susceptance of a capacitive diaphragm, b_C, is approximately given by

$$b_C \approx (4b/\lambda_g) \ln \operatorname{cosec} (\pi d_C/2b)$$

where d_C = width of the diaphragm gap
b = narrow dimension of the guide

A capacitive diaphragm markedly reduces the electric strength of the guide.

A *resonant diaphragm* (*iris* or *window*) is a thin metal sheet with a rectangular or oval opening (Fig. 2.21*c*), placed across the guide. It combines the features of an inductive and a capacitive diaphragm. The size of the opening in a resonant diaphragm can be chosen such that it will not affect the propagation of the TE_{10} mode down the guide, that is, it will have a zero susceptance at the specified resonant frequency. The equivalent circuit of a resonant window includes a parallel resonant circuit which shunts the guide. The resonant frequency of the diaphragm is approximately found to be such that the transmission line equivalent to the guide and the opening in the diaphragm have the same characteristic impedance as given by Eq. (2.1)

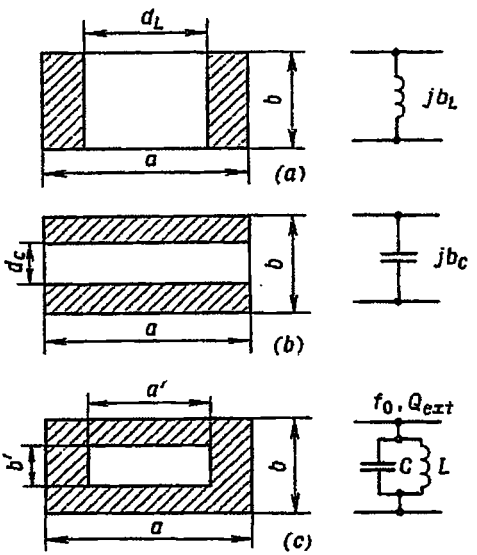


Fig. 2.21. Diaphragms in rectangular guide

$$\frac{b}{a\,[1-(\lambda_0/2a)^2]^{1/2}} = \frac{b'}{a'\,[1-(\lambda_0/2a')^2]^{1/2}}$$

It can be shown that the resonant wavelength λ_0 thus chosen can be obtained with a multitude of diaphragms varying in the size of the opening from a narrow slot $\lambda_0/2$ long to the full cross-section of the guide. These resonant diaphragms differ in their external Q, that is the Q of an equivalent resonant LC-circuit

$$Q_{ext} = z_c/2\ (L/C)^{1/2}$$

with allowance for the effect of the line matched at both ends and shunting this resonant circuit.

Inductive posts. An *inductive post* (Fig. 2.22*a*) is a round metal conductor installed in a rectangular guide in the direction of electric lines of force, E, and reaching as far as the broad faces of the guide at both ends. The equivalent circuit of an inductive post consists of a shunt inductance and two series capacitive reactances to

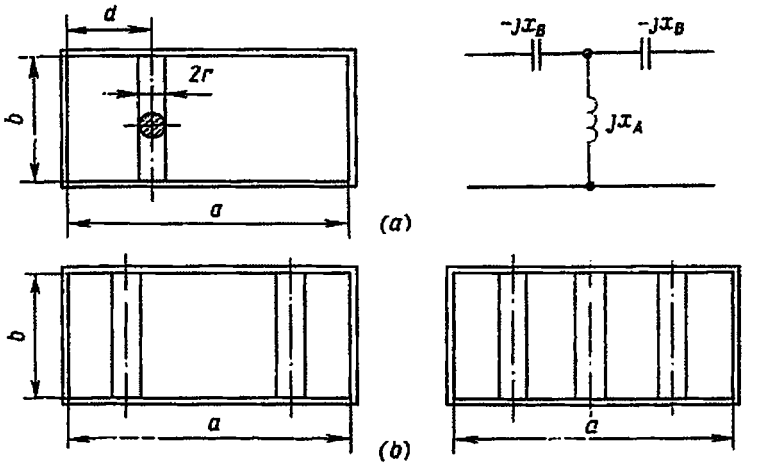


Fig. 2.22. Inductive post in rectangular guide

account for the finite thickness of the post. The element values can be found from equations and plots in reference books. Inductive posts do not impair the electric strength of the guide and are simple to make.

When it is desired to reduce the value of the shunt reactance x_A, an array of several inductive posts will usually be used; they are located in the cross-section of the guide as shown in Fig. 2.22*b*.

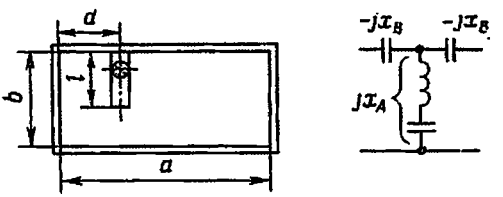


Fig. 2.23. Capacitive post in rectangular guide

Capacitive posts. A *capacitive post* (Fig. 2.23) is a round conductor installed in the direction of electric lines of force, and joined at one end to a broad face of the guide. The equivalent circuit of a capacitive post contains a series resonant LC-circuit shunting the transmission line. The capacitance of this resonant circuit accounts for the

concentration of the electric field at the free end of the post, and the inductance takes care of the currents flowing in the post.

At some length of the post, close to a quarter of the resonant wavelength, the admittance of the series resonant circuit turns to infinity, and the waveguide is shorted. Shorter posts show a capacitive admittance. At post lengths in excess of the resonant one, the admittance is inductive in its effect. The series capacitive reactances in the equivalent circuit account for the finite thickness of the post. At small post diameters, these reactances are small and may be neglected.

Capacitive posts are mainly used as adjustable reactive elements introduced into a guide through threaded holes in a broad face of the guide. For this reason they are often called capacitive tuning screws. Unfortunately, capacitive posts markedly impair the electric strength of the guides, which is the reason why they are not used in high-power guide systems.

2.8. Mismatch in a Microwave Transmission System with Many Discontinuities

When he sets out to design a microwave transmission system, the designer does not know in advance either the true values of the parameters that characterize each discontinuity or its exact location. Therefore, the input TWR is predicted statistically. Ideally, no one element in a microwave transmission system should introduce a mismatch. Actually, however, any kth element (where $k = 1, 2, 3, \ldots$) does have a reflection coefficient due to imperfections in manufacture and inaccuracy in alignment

$$\dot{\rho}_k = \rho_k \exp(j\varphi_k)$$

whose magnitude and phase vary in a random manner from specimen to specimen of the same element type. The spread in the magnitude of the reflection coefficient is usually stated in terms of the variance σ_i^2 derived from analysis of a large number M_{max} of identical discontinuities

$$\sigma_i^2 = \frac{1}{M_{max} - 1} \sum_{m=1}^{M_{max}} \rho_{im}^2$$

The equivalent circuit of a microwave transmission system containing random discontinuities is shown in Fig. 2.24. Assuming that the individual reflection coefficients of discontinuities are small ($\rho_k \ll 1$), the incident voltage wave in the system may be deemed constant. Using Eq. (1.17) repeatedly, we can combine the individual or partial reflection coefficients to obtain the overall re-

flection coefficient referred (or transferred) to the input of the system:

$$\dot{\rho} \approx \sum_{k=1}^{N} \rho_k \exp\left[-j\left(2\beta l_k - \varphi_k\right)\right] = \rho_x + j\rho_y \qquad (2.2)$$

Note that the phase angles of the individual terms are normally distributed random variables in the range from zero to 2π. Therefore

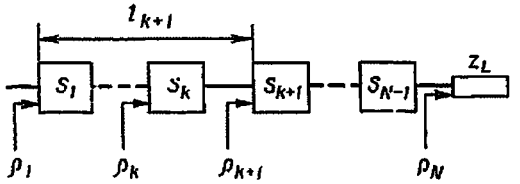


Fig. 2.24. Microwave transmission system with many discontinuities

the mean values of the rectangular components ρ_x and ρ_y in the input reflection coefficient are zero, and the components themselves obey the Gaussian distribution

$$w(\rho_x) = \frac{1}{(2\pi)^{1/2}\sigma_x} \exp\left(-\rho_x^2/2\sigma_x^2\right)$$

and

$$w(\rho_y) = \frac{1}{(2\pi)^{1/2}\sigma_y} \exp\left(-\rho_y^2/2\sigma_y^2\right)$$

where the variances σ_x^2 and σ_y^2 are equal to the sum of the variances of the real and imaginary parts of the individual components defined by Eq. (2.2):

$$\sigma_x^2 = \sigma_y^2 = \frac{1}{2}\sum_{k=1}^{N} \sigma_k^2 \equiv \sigma^2 \qquad (2.3)$$

In evaluating the quality of match for a microwave transmission system, what matters is solely the magnitude of the reflection coefficient for which the probability density function has a Rayleigh distribution*

$$w(\rho) = (\rho/\sigma^2) \exp\left(-\rho^2/2\sigma^2\right)$$

shown in the plot of Fig. 2.25. The parameter of the Rayleigh distribution is the most probable value of $\rho_m = \sigma$ corresponding to the peak on the distribution curve.

---

* As will be recalled, this is the distribution of the magnitude of a two-dimensional vector whose projections on the coordinate axes are normally distributed.

If we know the probability distribution, we can readily compute the probability that the magnitude of the reflection coefficient at the input of the microwave transmission system will be smaller than a certain predetermined maximum value, ρ_{max}. On integrating $w(\rho)$ with respect to ρ between the limits 0 and ρ_{max}, we obtain

$$\Phi(\rho < \rho_{max}) = 1 - \exp(-\rho_{max}^2/2\sigma^2)$$

For example, at a probability of $\Phi = 0.9$, the respective value of the maximum reflection coefficient will be $\rho_{max} = 2.14\sigma$, which leads to the following value of the TWR:

$$K_{0.9} = (1 - 2.14\sigma)/(1 + 2.14\sigma) \qquad (2.4)$$

The values of σ in the above equation can be found from Eq. (2.3).

The TWR found by Eq. (2.4) will be exceeded in 90% of the various microwave transmission systems designed and built from specified elements. With a poor design for which the probability is 10%, K is less than $K_{0.9}$, and the microwave transmission system must be re-configured, using other specimens of the same elements or varying their relative positions. Equations (2.4) and (2.3) also offer an opportunity to formulate the requirement for the quality of match between the individual components of the system. Note that the prediction of the TWR holds only for one operating frequency. If one wishes to appraise the quality of match for a transmission system operating over a bandwith Δf, the expected TWR must be computed at several frequencies differing by an amount such that the predicted value of the TWR at one will be statistically independent of that at any other frequency. The probability that the reflection coefficient at M frequencies will be smaller than ρ_{max} can be found by the following equation:

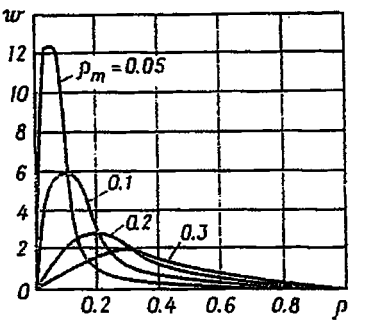


Fig. 2.25. Rayleigh distribution of the magnitude of the reflection coefficient in a microwave transmission system with discontinuities

$$\Phi_M(\rho < \rho_{max}) = [1 - \exp(-\rho_{max}^2/2\sigma^2]^M$$

where the term to the Mth power is the probability of one trial. The required number of trials, or estimates of the TWR over the bandwidth is given by the following empirical equation:

$$M = (\Delta f/f_0)(2L_{av}/\lambda_g)$$

where L_{av} = average spacing between the discontinuities
f_0 = centre frequency
λ_g = guide wavelength

If the relative bandwidth $\Delta f/f_0$ is not over 10%; then at L_{av} equal to from 5 to 7 guide wavelengths it will suffice to estimate the TWR at only the mid-band frequency.

Chapter Three

Microwave Multiport Junctions

3.1. Basic Definitions

The concept of a microwave multiport junction is crucial in an analysis of ramified microwave transmission systems.

A *microwave multiport junction* refers to a structure made up of conductors, dielectrics and other microwave components, having several points of access, called ports*, in the form of transverse planes on transmission lines, each capable of sustaining a specified wave mode. The ports of multiport junctions are called *terminal* or *reference planes*. Their locations are chosen such that the non-propagating modes belonging to the internal electromagnetic field will be negligible at those planes. This eliminates the likely transfer of energy between the multiport junction and the remainder of the system in any other way except by the specified mode in each line (or lead).

For many types of transmission line (hollow or surface waveguides) ports cannot be identified explicitly. So, when we speak of a microwave N-port junction, we mean a device with N incoming lines or, more rigorously, with N modes in all incoming transmission lines.

We will be mainly concerned with *passive linear multiport junctions*. The *property of passivity* implies the absence of amplification or generation of microwave power within the junction, or mathematically

$$P_{loss} \geqslant 0$$

where P_{loss} is the power dissipated by the multiport with any excitation at its ports.

The *property of linearity* implies that the external characteristics of a multiport junction are independent of the input microwave power level. However, this level must remain within certain limits, say, below the electric strength of the dielectrics used. Electric oscillations in linear multiport junctions are described by linear differential equations, while the techniques of linear algebra will suffice for analysis of steady-state conditions.

---

* At lower frequencies, a port is often called a terminal pair.—*Translator's note.*

Microwave circuit theory mainly deals with the external characteristics of multiport junctions, which relate the conditions that exist at their ports. The pattern and behaviour of the internal electromagnetic field are less important because these matters fall in the domain of applied electrodynamics. The external characteristics of linear multiport junctions are mainly described and analysed by recourse to the matrix technique of linear algebra — a technique best suited to the capabilities of computers.

The *matrices of a multiport junction* define the relations that exist between its input and output ports. The conditions existing at the

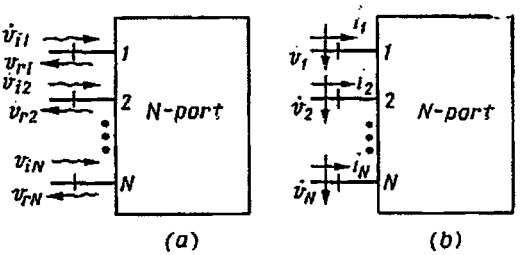


Fig. 3.1. Notation for a microwave multiport junction: *a*, the wave approach; *b*, the classical approach

reference planes of a multiport junction can be described in terms of the normalized incident and reflected voltage waves (the *wave approach*) or in terms of total normalized voltages and currents (the *classical approach* similar to that adopted in low-frequency circuit theory).

With the wave approach, let us agree, for each port of an arbitrary N-port junction, to call *incident* the normalized voltage wave $\dot{v}_{i\,m}$ (having the dimensions of $W^{1/2}$), propagating towards the multiport junction, and *reflected* (or scattered) the normalized voltage wave $\dot{v}_{r\,m}$ (having the dimensions of $W^{1/2}$) propagating away from the multiport junction (Fig. 3.1*a*). On any line, the incident and reflected waves are normalized according to Eq. (1.5). Owing to normalization, each incoming line is assigned in advance a dimensionless characteristic impedanace of unity, although the lines may actually differ from one another and operate in different modes.

With the classical approach, the conditions existing at each port of a multiport junction are specified in terms of the normalized voltages $\dot{v}_m$ and normalized currents $\dot{i}_m$ entering the multiport junction (Fig. 3.1*b*). Owing to normalization, the normalized voltages and currents have the same dimensions of $W^{1/2}$. According to Eqs. (1.9), (1.10) and (1.12), there is a simple connection between the wave and the classical descriptions of the conditions existing at

the ports of a multiport junction, namely

$$\left.\begin{aligned} \dot{v}_m &= \dot{v}_{im} + \dot{v}_{rm} \\ \dot{i}_m &= \dot{v}_{im} - \dot{v}_{rm} \end{aligned}\right\} \text{ for } \dot{v}_{im}/\dot{i}_{im} \equiv z_{cm} = 1 \tag{3.1}$$

Therefore, the conditions at each port of a multiport junction can, in the most general case, be uniquely specified by giving any two complex quantities encountered in Eq. (3.1). As an example, for the mth port these quantities may be $\dot{v}_{i\,m}$ and $\dot{v}_{r\,m}$ or $\dot{v}_m$ and $\dot{i}_m$ or $\dot{v}_{i\,m}$ and $\dot{i}_m$ (a total of 12 combinations). If we choose at each port of an N-port junction one of the complex quantities appearing in Eq. (3.1) as a component of the driving function (excitation or stimulus), and some other complex quantity as a component of the response to this excitation, we may form as many N-ary *excitation vectors* $\mathbf{f}\rangle$ as there are N ports, and as many N-ary *response vectors* $\mathbf{r}\rangle$, such as

$$\mathbf{f}\rangle \equiv \begin{bmatrix} \dot{f}_1 \\ \dot{f}_2 \\ \cdot \\ \cdot \\ \cdot \\ \dot{f}_N \end{bmatrix} = \begin{bmatrix} \dot{v}_1 \\ \dot{v}_2 \\ \cdot \\ \cdot \\ \cdot \\ \dot{v}_{im} \end{bmatrix}, \qquad \mathbf{r}\rangle \equiv \begin{bmatrix} \dot{r}_1 \\ \dot{r}_2 \\ \cdot \\ \cdot \\ \cdot \\ \dot{r}_N \end{bmatrix} = \begin{bmatrix} \dot{i}_1 \\ \dot{i}_2 \\ \cdot \\ \cdot \\ \cdot \\ \dot{i}_{rN} \end{bmatrix}.$$

Here and elsewhere, $\mathbf{a}\rangle$ denotes, as proposed by Dirac, a shorthand notation for a column matrix with elements $\dot{a}_1, \dot{a}_2, \ldots, \dot{a}_N$. Owing to the linearity of the multiport junction, the components of any response column matrix should be linearly connected to the components of the respective excitation column matrix

$$\begin{aligned} \dot{r}_1 &= \dot{t}_{11}\dot{f}_1 + \dot{t}_{12}\dot{f}_2 + \ldots + \dot{t}_{1N}\dot{f}_N \\ \dot{r}_2 &= \dot{t}_{21}\dot{f}_1 + \dot{t}_{22}\dot{f}_2 + \ldots + \dot{t}_{2N}\dot{f}_N \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \dot{r}_N &= \dot{t}_{N1}\dot{f}_1 + \dot{t}_{N2}\dot{f}_2 + \ldots + \dot{t}_{NN}\dot{f}_N \end{aligned}$$

which, in matrix notation, have the form

$$\begin{bmatrix} \dot{r}_1 \\ \dot{r}_2 \\ \cdots \\ \dot{r}_N \end{bmatrix} = \begin{bmatrix} \dot{t}_{11} & \dot{t}_{12} & \ldots & \dot{t}_{1N} \\ \dot{t}_{21} & \dot{t}_{22} & \ldots & \dot{t}_{2N} \\ \cdots & \cdots & \cdots & \cdots \\ \dot{t}_{N1} & \dot{t}_{N2} & \ldots & \dot{t}_{NN} \end{bmatrix} \begin{bmatrix} \dot{f}_1 \\ \dot{f}_2 \\ \cdots \\ \dot{f}_N \end{bmatrix}.$$

such that the square T matrix with elements $\dot{t}_{mn}$ is a complete external characteristic of the N-port junction in the sense that it permits the designer to determine the response of the multiport junction in the selected form of r⟩ to any excitation in the chosen form f⟩.

Because there may be 12 combinations of the quantities defining the excitation and response for each port, we may, at least in principle, characterize one and the same N-port junction by 12^N distinct T matrices. However, among this multitude, only very few matrices have a well-defined physical significance. Basically, three types of matrices are used to characterize multiports, namely: scattering matrices, impedance matrices, and admittance matrices.

3.2. The Scattering Matrix

In microwave circuit theory, it is most common to choose the excitation vector for an N-port junction as a set of N incident waves and the response vector as a set of N reflected waves, that is,

$$\mathrm{f}\rangle \equiv \mathrm{v}_i\rangle = \begin{bmatrix} \dot{v}_{i1} \\ \dot{v}_{i2} \\ \cdots \\ \dot{v}_{iN} \end{bmatrix}, \qquad \mathrm{r}\rangle \equiv \mathrm{v}_r\rangle = \begin{bmatrix} \dot{v}_{r1} \\ \dot{v}_{r2} \\ \cdots \\ \dot{v}_{rN} \end{bmatrix}$$

The excitation and response vectors thus defined are connected by the *scattering matrix* S. The set of linear equations characterizing the scattering matrix have the form

$$\begin{aligned}
\dot{v}_{r1} &= \dot{s}_{11}\dot{v}_{i1} + \dot{s}_{12}\dot{v}_{i2} + \cdots + \dot{s}_{NN}\dot{v}_{1N} \\
\dot{v}_{r2} &= \dot{s}_{21}\dot{v}_{i1} + \dot{s}_{22}\dot{v}_{i2} + \cdots + \dot{s}_{2N}\dot{v}_{1N} \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
\dot{v}_{rN} &= \dot{s}_{N1}\dot{v}_{i1} + \dot{s}_{N2}\dot{v}_{i2} + \cdots + \dot{s}_{NN}\dot{v}_{1N}
\end{aligned}$$

In matrix notation the above set of equations takes the form

$$\begin{bmatrix} \dot{v}_{r1} \\ \dot{v}_{r2} \\ \cdots \\ \dot{v}_{rN} \end{bmatrix} = \begin{bmatrix} \dot{s}_{11}\dot{s}_{12} & \cdots & \dot{s}_{1N} \\ \dot{s}_{21}\dot{s}_{22} & \cdots & \dot{s}_{2N} \\ \cdots\cdots\cdots & & \\ \dot{s}_{N1}\dot{s}_{N2} & \cdots & \dot{s}_{NN} \end{bmatrix} \begin{bmatrix} \dot{v}_{i1} \\ \dot{v}_{i2} \\ \cdots \\ \dot{v}_{iN} \end{bmatrix} \tag{3.2}$$

or, in a more compact form (which will be used in the subsequent discussion),

$$\mathbf{v}_r\rangle = \mathbf{S}\mathbf{v}_i\rangle \tag{3.2a}$$

The square matrix S in (3.2) and (3.2a) has the sense of a mathematical operator defining the rule by which the excitation vector $\mathbf{v}_i\rangle$ can be transformed into the response vector $\mathbf{v}_r\rangle$. Equations (3.2) and (3.2a) bear a close resemblance to the relation

$$\dot{v}_r = \dot{\rho}\dot{v}_i$$

for one-ports, where the reflection coefficient $\dot{\rho}$ plays the role of the scattering matrix. This is reasonable because a one-port is in effect the simplest form of a multiport junction.

In order to determine the elements of the scattering matrix (or the elements of any other matrices of a multiport junction for that matter), we must analyse by computation or experiment the behaviour of the multiport junction under a set of test conditions. For a general N-port junction, the total number of such sets of test conditions should be no less than the number N where N is the order of the matrix equal to the number of ports on the junction.

There are the *simplest test conditions* which directly lead to the various elements of the matrix. For the scattering matrix, such conditions involve the excitation of the multiport junction under test by incident waves applied in turn to each port. On all other lines, except the excited one, there should be no incident waves. Therefore, outside the multiport junction these lines must be terminated in nonreflecting absorbers, or matched loads. Thus, each simplest set of test conditions is obtained by connecting the source of the incident wave to one of its ports and matched loads to all other ports.

Referring to Eqs. (3.2), it is seen that if one of the incident voltage waves is nonzero, it is an easy matter to find the respective column in the scattering matrix. Obviously, this column bears the same ordinal number as the excited port, and the elements of this column are numerically equal to the ratio of the normalized voltage waves reflected from the multiport to the only normalized incident voltage wave:

$$\dot{s}_{mn} = \frac{\dot{v}_{rm}}{\dot{v}_{im}}\Bigg|_{\dot{v}_{iq}=0,\ q=1,\ 2,\ \ldots,\ N,\ q\neq n} \tag{3.3}$$

The complex quantities $\dot{v}_{rm}$ and $\dot{v}_{im}$ in (3.3) must be unambiguously referenced to the reference planes of the multiport junction. In other words, these quantities must be calculated or measured at the preselected reference planes on each incoming line of the multiport.

As follows from (3.3), the elements of the scattering matrix (the S parameters or the scattering coefficients) are dimensionless and have

each a well-defined physical interpretation. The off-diagonal elements of the S matrix are the normalized-voltage wave transmission coefficients between each pair of ports with all the other ports terminated in matched loads. Each of the diagonal elements for which $m = n$ is the reflection coefficient of that port, with all the other ports terminated in matched loads. It is useful to remember that the first letter m in the subscript of each scattering coefficient $\dot{s}_{mn}$ denotes both the matrix row and the matched port to which power is transmitted. The second letter n denotes the column of the matrix and the port to which excitation is applied.

From the definition of the scattering coefficients, (3.3), it follows that for passive multiport junctions (that is, multiports incapable of power amplification), the magnitudes of the transmission and reflection coefficients, $|\dot{s}_{mn}|$, cannot exceed unity. This implies that the scattering matrix must exist for any passive, linear multiport junction.

The scattering matrix (as any other matrix, in fact) can describe the behaviour of a multiport junction at some particular frequency only. In cases involving a band of frequencies, the elements of the scattering and other matrices become complex-valued functions of frequency.

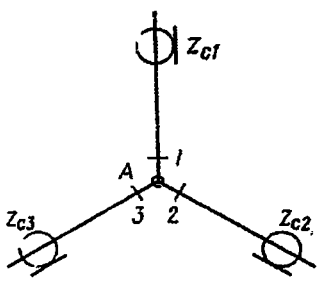


Fig. 3.2. Parallel (shunt) junction of transmission lines

Example 3.1. The scattering matrix of a coaxial junction (Fig. 3.2). Let three coaxial cables operating in the TEM mode and differing in characteristic impedance, $Z_{c\,i}$, be connected in parallel at point A. If the cable diameters are small in comparison with the wavelength, we may neglect the effect of the higher-order modes in the vicinity of point A and assume that the non-normalized voltages on the lines excited in any arbitrary manner are the same at that point, that is, $\dot{V}_1 = \dot{V}_2 = \dot{V}_3 = \dot{V}_A$. Let the reference plane on each line be located at an infinitesimally short distance from point A (this is a valid assumption because we have also assumed that there are no higher-order modes in the vicinity of point A).

Let, in the first set of test conditions, the non-normalized incident voltage wave $\dot{V}_{i1}$ enter the junction via port *1*. If ports *2* and *3* are terminated in non-reflecting absorbers (that is, resistors for which the characteristic impedances are Z_{c2} and Z_{c3}), the equivalent load on the first line at point A will be a parallel combination of Z_{c2} and Z_{c3}, equal to

$$R_L = Z_{c2}Z_{c3}/(Z_{c2} + Z_{c3})$$

In accordance with Eq. (3.3), the reflection coefficient for this load, that is, the scattering coefficient $\dot{s}_{11}$, will be

$$\dot{s}_{11}=\frac{R_L-Z_{c1}}{R_L+Z_{c1}}=\frac{Z_{c2}Z_{c3}-Z_{c1}Z_{c2}-Z_{c1}Z_{c3}}{Z_{c1}Z_{c2}+Z_{c1}Z_{c3}+Z_{c2}Z_{c3}}$$

The total non-normalized voltage at point A will be

$$\dot{V}_A=V_{i1}(1+\dot{s}_{11})=\dot{V}_{i1}\frac{2Z_{c2}Z_{c3}}{Z_{c1}Z_{c3}+Z_{c1}Z_{c3}+Z_{c2}Z_{c3}}=\dot{V}_{r2}=\dot{V}_{r3}$$

Since, in agreement with Eq. (1.15), the normalized incident voltage wave on line *1* is

$$\dot{v}_{i1}=\dot{V}_{i1}/\sqrt{Z_{c1}}$$

and the normalized scattered voltage waves on lines *2* and *3* respectively are

$$\dot{v}_{r2}=\dot{V}_A/\sqrt{Z_{c2}}$$

and

$$\dot{v}_{r3}=\dot{V}_A/\sqrt{Z_{c3}}$$

the following expressions can be written for the scattering coefficients $\dot{s}_{21}$ and $\dot{s}_{31}$, according to Eq. (3.3):

$$\dot{s}_{21}=\dot{v}_{r2}/\dot{v}_{i1}=\frac{2Z_{c3}(Z_{c1}Z_{c2})^{1/2}}{Z_{c1}Z_{c3}+Z_{c1}Z_{c3}+Z_{c2}Z_{c3}}$$

$$\dot{s}_{31}=\dot{v}_{r3}/\dot{v}_{i1}=\frac{2Z_{c3}(Z_{c1}Z_{c3})^{1/2}}{Z_{c1}Z_{c2}+Z_{c1}Z_{c3}+Z_{c2}Z_{c3}}$$

The analysis procedure is the same when the junction is excited via port *2* or *3*, with all the other ports terminated in matched loads. Instead of actual tests, however, the same results can be obtained on just replacing the subscripts in the above equations as follows: 1 by 2, 2 by 3, and 3 by 1 when the excitation is applied to port *2*, and 1 by 3, 3 by 2, and 2 by 1, when the excitation is applied to port *3*.

Exciting each port of a multiport junction in turn and terminating all the other ports in matched loads is not always the shortest route to deriving the scattering matrix. In some cases, other forms of tests may prove advantageous. In the general case, excitation of an N-port by M incident waves can be described by a rectangular excitation matrix

$$\mathbf{V}_i=\begin{bmatrix}\dot{v}_{i11} & \dot{v}_{i12} & \dots & \dot{v}_{i1M}\\ \dot{v}_{i21} & \dot{v}_{i22} & \dots & \dot{v}_{i2M}\\ \dots & \dots & \dots & \dots\\ \dot{v}_{iN1} & \dot{v}_{iN2} & \dots & \dot{v}_{iNM}\end{bmatrix} \tag{3.4}$$

In this matrix, the columns of dimension N contain the incident voltages in each set of test conditions. The respective responses of the N-port junction in the form of M sets of N-ary scattered (reflected) wave vectors form a rectangular response matrix

$$\mathbf{V}_r = \begin{bmatrix} \dot{v}_{r11} & \dot{v}_{r12} & \dots & \dot{v}_{r1M} \\ \dot{v}_{r21} & \dot{v}_{r22} & \dots & \dot{v}_{r2M} \\ \dots & \dots & \dots & \dots \\ \dot{v}_{rN1} & \dot{v}_{rN2} & \dots & \dot{v}_{rNM} \end{bmatrix} \tag{3.5}$$

For the same N-port junction, the excitation and response matrices are interconnected by a scattering matrix of the form

$$\begin{bmatrix} \dot{v}_{r11} & \dots & \dot{v}_{r1M} \\ \dots & \dots & \dots \\ \dot{v}_{rN1} & \dots & \dot{v}_{rNM} \end{bmatrix} = \begin{bmatrix} \dot{s}_{11} & \dots & \dot{s}_{1N} \\ \dots & \dots & \dots \\ \dot{s}_{N1} & \dots & \dot{s}_{NN} \end{bmatrix} \begin{bmatrix} \dot{v}_{i11} & \dots & \dot{v}_{i1M} \\ \dots & \dots & \dots \\ \dot{v}_{iN1} & \dots & \dot{v}_{iNM} \end{bmatrix} \tag{3.6}$$

If the excitation matrix $\mathbf{V}_i$ has been specified in advance and the corresponding response matrix $\mathbf{V}_r$ has been found by computation or experiment, Eq. (3.6) may be used as a matrix equation in the unknown scattering matrix S. In the simplest case, in order to solve Eq. (3.6) for the unknown scattering matrix S, it will suffice to select just a number N of linearly independent excitations, that is, to consider the case where $M = N$ for which the excitation matrix will be a square, nonsingular one. Then for the excitation matrix $\mathbf{V}_i$ there will exist one and only one inverse matrix, $\mathbf{V}_i^{-1}$. Post-multiplying Eq. (3.6) by $\mathbf{V}_i^{-1}$ and recalling that

$$\mathbf{V}_i\mathbf{V}_i^{-1} = \mathbf{V}_i^{-1}\mathbf{V}_i = \mathbf{E}$$

(where E is the square unit matrix of order N in which the elements on the main diagonal are ones and all other elements are zeros), we may write

$$\mathbf{S} = \mathbf{V}_r\mathbf{V}_i^{-1} \tag{3.7}$$

When the ports of a multiport junction are excited one at a time with a unity incident wave, the excitation matrix $\mathbf{V}_i$ reduces to an identity (unit) matrix. Then it immediately follows from Eq. (3.7) that all the scattering coefficients in such a case are numerically equal to the normalized scattered (reflected) voltage waves, forming the square response matrix $\mathbf{V}_r$. Precisely for this reason this case was taken up earlier in the discussion of the simplest tests.

The scattering coefficients can directly be found by any one of several experimental procedures. The simplest set of test equipment should include a microwave generator, an impedance meter or a slo-

tted line, a microwave power meter, assorted matched loads, and movable shorted stubs. Connecting these devices in various combinations to the N-port junction of interest, it is possible to observe its excitations and responses under several sets of test conditions, which is sufficient for the respective scattering matrix to be computed. Several types of microprocessor-based automatic and semiautomatic systems have been developed which can rapidly measure the magnitudes and phases of the scattering coefficients for any microwave circuit over a wide dynamic range and over the desired bandwidth.

In the general case, measurement of the scattering coefficients consists in applying a series of independent test excitations and storing the respective column matrices of incident and reflected waves in a memory unit. Then the column matrices are combined into rectangular matrices as defined by (3.4) and (3.5), and the matrix equation (3.6) is solved for the scattering coefficients. Because experimental records are always subject to random errors arising from imperfections in the test equipment and other sources, it is reasonable to make replicate measurements, with the number M of combinations of test excitations exceeding the number N of ports on the junction under test, so that the results of solving Eq. (3.6) could appropriately be averaged.

With M exceeding N, the excitation matrix $\mathbf{V}_i$ is a rectangular one, and no conventional inverse matrix exists for it. Therefore, Eq. (3.6) cannot be solved directly in this case. When M exceeds N, the relations given by Eq. (3.6) are an overdefined set of linear equations in the unknown scattering coefficients. It is reasonable to seek a solution for s_{mn} by the least squares method. It is proved in linear algebra that a solution accurate to within the standard deviaction is given by an expression similar to Eq. (3.7):

$$\mathbf{S} = \mathbf{V}_r \mathbf{V}_i^{\dagger}$$

where $\mathbf{V}_i^{\dagger}$ is the so-called pseudo-inverse (generalized inverse) matrix. In our case, the pseudo-inverse matrix has N columns with M elements in each, and can be found by applying, say, the following equation:

$$\mathbf{V}_i^{\dagger} = \mathbf{V}_{it}^{*} [\mathbf{V}_i \mathbf{V}_{it}^{*}]^{-1}$$

where the subscript t denotes the operation of transposition, and the asterisk (*) denotes taking the complex conjugate of all elements. Resort to pseudo-inversion complicates the computation of the scattering matrix only slightly as compared with Eq. (3.7). However, it markedly enhances the accuracy of the result because the random errors contained in the $\mathbf{V}_i$ and $\mathbf{V}_r$ matrices at $M > N$ are automatically averaged.

3.3. The Impedance and Admittance Matrices

In determining the impedance and admittance matrices, one uses the classical approach to the description of the conditions existing at the ports of a multiport junction in terms of total normalized voltages and currents. This yields a matrix description of microwave multiport junctions almost identical to that adopted in the theory of low-frequency circuits. The principal distinction consists in that the usual voltages and currents are replaced by their normalized equivalents with the dimensions of $(\text{watt})^{1/2}$.

To begin with, let us define the *impedance matrix* Z. Let the excitation applied to an N-port junction be a set of N normalized currents and the response be a set of N normalized voltages, that is,

$$|\mathbf{f}\rangle \equiv |\mathbf{i}\rangle = \begin{bmatrix} \dot{i}_1 \\ \dot{i}_2 \\ \cdots \\ \dot{i}_N \end{bmatrix}, \qquad |\mathbf{r}\rangle \equiv |\mathbf{v}\rangle = \begin{bmatrix} \dot{v}_1 \\ \dot{v}_2 \\ \cdots \\ \dot{v}_N \end{bmatrix}$$

The excitation and response vectors, or column matrices, defined above are connected by a normalized impedance matrix, Z. In matrix notation, the set of linear algebraic equations describing the Z matrix has the form

$$|\mathbf{v}\rangle = \mathbf{Z}|\mathbf{i}\rangle \tag{3.8}$$

or, in a detailed notation,

$$\begin{bmatrix} \dot{v}_1 \\ \dot{v}_2 \\ \cdots \\ \dot{v}_N \end{bmatrix} = \begin{bmatrix} \dot{z}_{11} & \dot{z}_{12} & \cdots & \dot{z}_{1N} \\ \dot{z}_{21} & \dot{z}_{22} & \cdots & \dot{z}_{2N} \\ \cdots & \cdots & \cdots & \cdots \\ \dot{z}_{N1} & \dot{z}_{N2} & \cdots & \dot{z}_{NN} \end{bmatrix} \begin{bmatrix} \dot{i}_1 \\ \dot{i}_2 \\ \cdots \\ \dot{i}_N \end{bmatrix} \tag{3.8a}$$

Equation (3.8) resembles very closely the conventional Ohm's law for a two-terminal circuit in the form

$$\dot{v} = \dot{z}\dot{i}$$

where the input impedance $\dot{z}$ plays the role of the impedance matrix Z. Using Eqs. (3.8a) and running an open-circuit test mentally at all ports except the excited one, it is an easy matter to identify the physical significance of the elements of the Z matrix. With the nth port excited by an ideal current source of normalized value $\dot{i}_n$ and with all the other ports open-circuited, all elements in the nth

column of the Z matrix may be written as

$$\dot{z}_{mn} = \frac{\dot{v}_m}{\dot{i}_n}\bigg|_{\dot{i}_q \equiv 0,\ q=1,\ 2,\ \ldots,\ N;\ q \neq n} \tag{3.9}$$

It should always be remembered that the complex normalized voltages and the driving normalized currents in Eq. (3.9) must be determined at the reference planes of the multiport junction.

The off-diagonal elements of the Z matrix are the *mutual impedances* between ports m and n on the junction. The first letter, m, in the subscript of $\dot{z}_{mn}$ indicates the row of the matrix and, at the same time, the open-circuited port where the corresponding response is measured as the normalized open-circuit voltage. The second letter, n, in the subscript denotes the column of the Z matrix and, at the same time, the port to which the excitation is applied in the form of a normalized current.

The diagonal elements of the Z matrix in the case where m is equal to n in Eq. (3.9) are each the *self-impedance* of the respective port, with all the other ports open-circuited. Because the normalized currents and voltages have the same dimensions of $(\text{watt})^{1/2}$, all elements of the impedance matrix are dimensionless.

Now let us define the normalized *admittance matrix* Y. Here, the excitation applied to the ports of an N-port junction makes up a set of N normalized voltages, and the respective response is a set of N normalized currents. In matrix notation, the set of linear algebraic equations describing the Y matrix has the form

$$|\dot{i}\rangle = \mathbf{Y}|\dot{v}\rangle$$

or, in detailed matrix notation,

$$\begin{bmatrix} \dot{i}_1 \\ \dot{i}_2 \\ \cdots \\ \dot{i}_N \end{bmatrix} = \begin{bmatrix} \dot{y}_{11} & \dot{y}_{12} & \cdots & \dot{y}_{1N} \\ \dot{y}_{21} & \dot{y}_{22} & \cdots & \dot{y}_{2N} \\ \cdots & \cdots & \cdots & \cdots \\ \dot{y}_{N1} & \dot{y}_{N2} & \cdots & \dot{y}_{NN} \end{bmatrix} \begin{bmatrix} \dot{v}_1 \\ \dot{v}_2 \\ \cdots \\ \dot{v}_N \end{bmatrix}$$

Using the above set of equations and short-circuiting mentally all ports except the driven one, we obtain the following expression for the elements of the admittance matrix:

$$\dot{y}_{mn} = \frac{\dot{i}_m}{\dot{v}_n}\bigg|_{\dot{v}_q \equiv 0,\ q=1,\ 2,\ \ldots,\ N;\ q \neq n}$$

The off-diagonal elements of the Y matrix are the complex *mutual admittances* defined as the ratios of the output normalized short-circuit currents to the normalized voltage at the driven port. The

diagonal elements of the Y matrix are each the *self-admittance* of the respective port, with all the other ports short-circuited. As with the impedance matrix, the normalized currents and voltages must be determined at the pre-specified reference planes. As with the Z matrix, the elements of the Y matrix are dimensionless.

From comparison of the Z and Y matrices of the same N-port junction it is an easy matter to conclude that they are mutually inverse:

$$\mathbf{Z} = \mathbf{Y}^{-1}, \quad \mathbf{Y} = \mathbf{Z}^{-1}$$

The impedance and admittance matrices are most often used in the analysis and synthesis of multi-element antennas in order to establish how the individual radiators affect one another's performance.

Note that for some passive multiport junctions either the impedance matrix or the admittance matrix or both may prove indeterminate (containing infinitely large elements).

3.4. Relations Between the Matrices of a Multiport Junction

Any parameter matrix of a multiport junction gives the complete external characteristic of that junction. Which matrix to use depends on convenience in element calculation or the experimental technique adopted. All matrices of one and the same multiport junction are related to one another in an inambiguous way. Therefore, if we know one matrix, we can readily deduce any other. As already noted, the relation is the simplest between the impedance and admittance matrices — one is the inverse of the other.

To establish the connection between the scattering matrix and the impedance (or admittance) matrix and generally between any two arbitrary matrices of a multiport junction, one needs to use Eq. (3.1) which connects the incident and reflected voltage waves on the input transmission lines, on the one hand, and the normalized voltages and currents, on the other. Recalling that Eqs. (3.1) hold for all input lines of a multiport junction, they can be re-cast as the following combined matrix:

$$2\mathbf{v}_i\rangle = \mathbf{v}\rangle + \mathbf{i}\rangle$$

or (3.10)

$$2\mathbf{v}_r\rangle = \mathbf{v}\rangle - \mathbf{i}\rangle$$

Substituting the column matrices $\mathbf{v}_i\rangle$ and $\mathbf{v}_r\rangle$ from (3.10) into (3.2) which defines the scattering matrix, we obtain

$$(\mathbf{v}\rangle - \mathbf{i}\rangle) = \mathbf{S}\,(\mathbf{v}\rangle + \mathbf{i}\rangle)$$

Collecting the terms which contain the factor $\mathbf{v}\rangle$ on the left-hand side of the above equation, we may write

$$(\mathbf{E} - \mathbf{S})\,\mathbf{v}\rangle = (\mathbf{E} + \mathbf{S})\,\mathbf{i}\rangle$$

where $\mathbf{E}$ is a unit matrix of order N. Premultiplying the above equation by the matrix $(\mathbf{E} - \mathbf{S})^{-1}$, we obtain

$$\mathbf{v}\rangle = (\mathbf{E} - \mathbf{S})^{-1}\,(\mathbf{E} + \mathbf{S})\,\mathbf{i}\rangle$$

On comparing it with Eqs. (3.8) which define the impedance matrix, we obtain the equation that defines the connection between the $\mathbf{Z}$ and $\mathbf{S}$ matrices:

$$\mathbf{Z} = (\mathbf{E} - \mathbf{S})^{-1}\,(\mathbf{E} + \mathbf{S}) \tag{3.11}$$

From Eq. (3.11) it follows that the impedance matrix does not exist always — it becomes indeterminate when the determinant of the matrix to be inverted reduces to zero, that is, when

$$\det\,(\mathbf{E} - \mathbf{S}) = 0$$

Applying the same reasoning, we can deduce an expression connecting the admittance matrix and the scattering matrix:

$$\mathbf{Y} = (\mathbf{E} + \mathbf{S})^{-1}\,(\mathbf{E} - \mathbf{S}) \tag{3.12}$$

It follows from Eq. (3.12) that if

$$\det\,(\mathbf{E} + \mathbf{S}) = 0$$

the multiport junction in question will have no $\mathbf{Y}$ matrix. For example, it can be established by inspection that $\det\,(\mathbf{E} - \mathbf{S})$ and $\det\,(\mathbf{E} + \mathbf{S})$ reduce to zero for a 3-port in the form of a parallel junction of three transmission lines (see Fig. 3.2). This is an indication that the multiport junction in question has neither a $\mathbf{Z}$ nor a $\mathbf{Y}$ matrix.

Outwardly, the matrix relations (3.11) and (3.12) resemble the well-known relations for a transmission line, which connect the normalized impedance $\dot{z}$ (or the normalized admittance $\dot{y}$) and the reflection coefficient:

$$\dot{z} = (1 + \dot{\rho})/(1 - \dot{\rho})$$

$$\dot{y} = (1 - \dot{\rho})/(1 + \dot{\rho})$$

which is quite natural because the above expressions may be treated as a special case of (3.11) and (3.12), applicable to a one-port.

To sum up, when it is desired to establish the relation that connects any two normalized matrices of a multiport junction, it will suffice to use the sets of equations defining these matrices and relations of the form defined by (3.10) connecting the normalized incident and reflected voltage and current waves. For example, it can be shown

(we leave this as an exercise for the reader) that the scattering matrix can be expressed in terms of the impedance and admittance matrices as follows:

$$\mathbf{S} = (\mathbf{Z} - \mathbf{E})(\mathbf{Z} + \mathbf{E})^{-1} = (\mathbf{E} - \mathbf{Y})(\mathbf{E} + \mathbf{Y})^{-1} \qquad (3.13)$$

which is not unlike the relations applicable to transmission lines

$$\dot{\rho} = (\dot{z} - 1)/(\dot{z} + 1) = (1 - \dot{y})/(1 + \dot{y})$$

Note that the matrix factors in Eqs. (3.11) through (3.13) are commutative, which can readily be verified with the aid of identity transformations. This opportunity to change the order of cofactors simplifies computation in the analysis of microwave devices.

Dependence of the multiport matrices on the port numbering scheme. Any matrix of a multiport junction will be meaningful only if its ports are numbered in a certain definite order. A change in the numbering scheme leads to other matrices. Of course, the numerical values of the elements will remain unchanged, but the position of the elements in the matrix will be different.

To deduce appropriate design relations for a given N-port junction, let us set up a square renumbering matrix G by the following rule. Write $N - 1$ zeros in each row of the G matrix and a one in the cell bearing the new number of the port whose previous number was the same as that of a given row. For example, if port *3* becomes port *7*, a one should be written in cell *7* of the third row, and zeros should be written instead of all other elements in that same row. Using concrete examples, the reader is invited to verify that the renumbering matrix thus obtained is orthogonal, which means that it satisfies the relation

$$\mathbf{G}\mathbf{G}_t = \mathbf{E}$$

where E is the unit matrix of order N.

If we denote the incident and reflected voltage waves as $\mathbf{v}_i\rangle$ and $\mathbf{v}_r\rangle$, and as $\mathbf{v}'_i\rangle$ and $\mathbf{v}'_r\rangle$, using, respectively the old and the new numbering schemes, they will be connected by the renumbering matrix as follows:

$$\mathbf{v}_i\rangle = \mathbf{G}\mathbf{v}'_i\rangle$$

and

$$\mathbf{v}_r\rangle \mathbf{G}\mathbf{v}'_r\rangle$$

On substituting the $\mathbf{v}_i\rangle$ and $\mathbf{v}_r\rangle$ column matrices in the set of equations (3.2) defining the scattering matrix, we obtain

$$\mathbf{G}\mathbf{v}'_r\rangle = \mathbf{S}\mathbf{G}\mathbf{v}'_i\rangle$$

Premultiplying the above equality by the $\mathbf{G}^{-1}$ matrix (owing to the property of orthogonality, $\mathbf{G}^{-1} = \mathbf{G}_t$), we get

$$\mathbf{v}'_r\rangle \mathbf{G}_t\mathbf{S}\mathbf{G}\mathbf{v}'_i\rangle$$

Hence, the new scattering matrix S_G for the multiport junction with its ports re-numbered is connected to the original scattering matrix by a relation of the form

$$S_G = G_t S G \qquad (3.14)$$

which is a special case of the similarity transformation known from mathematics.

Reasoning as before, we can obtain expressions similar to Eqs. (3.14) for the impedance and admittance matrices of the multiport junction with its ports re-numbered:

$$Z_G = G_t Z G$$

and

$$Y_G = G_t Y G$$

Port renumbering offers a possibility to reduce the matrices of various devices to a standard format adopted for a particular class of multiport junctions.

Dependence of the multiport matrices on a change in the location of terminal planes. In some cases, a need arises to know how to rearrange the matrices of a multiport junction when its terminal planes have been shifted to a new location. The simplest case is the scattering matrix. When the terminal planes are moved away from the junction in question, further lagging phase shifts are introduced into the scattering coefficients because the signals have now a longer distance to travel. Also, the waves suffer an attenuation in the incoming lines, and the scattering coefficients become smaller in magnitude. As a result, any scattering coefficient, as reckoned from the new terminal planes, takes the form

$$\hat{s}_{mn} = \dot{s}_{mn} \exp(-j\dot{\gamma}_m l_m - j\dot{\gamma}_n l_n) \qquad (3.15)$$

where l_m and l_n are the increases in the length of the mth and nth incoming line, respectively, while $\dot{\gamma}_m = \beta_m - j\alpha_m$ and $\dot{\gamma}_n = \beta_n - j\alpha_n$ are the propagation constants of the same lines.

A change in the location of the terminal planes also leads to changes in the elements of the impedance and admittance matrices. However, it is not at all easy to deduce the new matrices $\hat{Z}$ and $\hat{Y}$ from the old matrices as is the case with the scattering matrix — this has to be done by transforming these matrices to the scattering matrix and back.

Normalized and non-normalized matrices of multiport junctions. Let us dwell in brief on the relation between the normalized and non-normalized impedance and admittance matrices. Non-normalized matrices are used on a limited scale in the analysis and synthesis of microwave devices in which the incoming lines sustain TEM modes. For such lines, the voltage and current normalization relations take

the form

$$\dot{v} = \dot{V}_m / \sqrt{Z_{cm}}$$

and

$$\dot{i}_m = \dot{I}_m \sqrt{Z_{cm}}$$

where m designates a particular port [see also Eqs. (1.15)]. For all the input lines of an N-port junction the following matrix normalizing relations hold:

$$\mathbf{v}\rangle = \{\mathbf{Z}_c\}^{-1/2}\,\mathbf{V}\rangle$$

and

$$\mathbf{i}\rangle = \{\mathbf{Z}_c\}^{1/2}\mathbf{I}\rangle$$

where $\{\mathbf{Z}_c\}^{1/2}$ is a diagonal matrix in which the diagonal elements are positive numbers $\sqrt{Z_{c\,m}}$ ($m = 1, 2, \ldots, N$). $\{\mathbf{Z}_c\}^{-1/2}$ is likewise a diagonal matrix in which the diagonal elements are equal to $1/\sqrt{Z_{c\,m}}$.

On substituting for v⟩ and i⟩ from the normalization equations into Eq. (3.8) and on solving this set of equations for the V column matrix, we obtain

$$\mathbf{V}\rangle = \{\mathbf{Z}_c\}^{1/2}\mathbf{Z}\{\mathbf{Z}_c\}^{1/2}\mathbf{I}\rangle$$

or

$$\mathbf{V}\rangle = \tilde{\mathbf{Z}}\mathbf{I}\rangle$$

Hence, the non-normalized voltage and current column matrices are connected by a square matrix of the form

$$\tilde{\mathbf{Z}} = \{\mathbf{Z}_c\}^{1/2}\mathbf{Z}\{\mathbf{Z}_c\}^{1/2}$$

which may be called the *non-normalized impedance matrix*. Its elements have the dimensions of ohms and are connected to the respective dimensionless elements $\dot{z}_{mn}$ of the normalized Z matrix by a relation of the form

$$\dot{Z}_{mn} = \dot{z}_{mn}\,(Z_{c\,m}Z_{c\,n})^{1/2}$$

Proceeding in a similar way, we can derive the *non-normalized admittance matrix*

$$\mathbf{I}\rangle = \tilde{\mathbf{Y}}\mathbf{V}\rangle$$

$$\tilde{\mathbf{Y}} = \{\mathbf{Z}_c\}^{-1/2}\,\mathbf{Y}\{\mathbf{Z}_c\}^{-1/2}$$

The elements of the $\tilde{\mathbf{Y}}$ matrix (which have the dimensions of siemens) are connected to the dimensionless elements of the normalized ad-

mittance matrix by a relation of the form

$$\dot{Y}_{mn} = \dot{y}_{mn}/(Z_{c\,m}Z_{c\,n})^{1/2}$$

The non-normalized matrices $\tilde{\mathbf{Z}}$ and $\tilde{\mathbf{Y}}$ find application in the theory of dipole and slotted antennas.

3.5. Ideal and Real Matrices of a Multiport Junction

The parameter matrices of multiport junctions may, by purpose and application, be classed into two broad groups.

Those in the first are *ideal matrices*. They serve to specify various microwave circuits and components proceeding from what they are to do in a microwave transmission system. The elements of these matrices are found by postulating the desired behaviour of a given system component in view of the objective it is to serve, rather than by analyzing the internal arrangement of a physical multiport junction.

For example, many microwave generators (say, magnetrons) need a protection against the detrimental effect of the waves reflected from the load. This job can be done by a two-port junction that would transmit microwave energy in one direction (from port *1* to port *2*) and block its passage in the reverse direction. Obviously, such a device, called an *isolator*, must have the following scattering matrix:

$$\mathbf{S} = \begin{array}{c} \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{cc} 1 & \quad 2 \end{array} \\ \begin{bmatrix} 0 & 0 \\ \exp(-j\varphi) & 0 \end{bmatrix} \end{array}$$

In specifying the above matrix, it is entirely immaterial to know how the isolator is constructed physically. In writing ideal matrices, one often ignores the frequency dependence of the elements and does not give the exact location of the terminal planes.

Those in the second group are *real matrices*. They present the results of calculations or experiments on real microwave circuits. Real matrices are always written, having in view a particular device, with the location of its terminal planes specified in advance and with allowance for the frequency behaviour of its elements in most cases.

For example, a particular specimen of the isolator may have the following real scattering matrix at the design frequency

$$\mathbf{S}_r = \begin{bmatrix} \dot{\rho}_1 & \dot{\alpha} \\ \dot{\tau} & \dot{\rho}_2 \end{bmatrix}$$

For a well designed and properly built isolator, the magnitudes of the scattering coefficients ρ_1, ρ_2 and α, are close to zero, and the mag-

nitude of the scattering coefficient τ is a little smaller than unity because of the inevitable internal losses. If measurements are made over a band of frequencies, the scattering coefficients will vary in magnitude, and their comparison with some limiting values specified in advance will establish the operating bandwidth of the isolator.

In setting up the ideal and real matrices of a multiport junction, a good deal depends on *a priori* knowledge about the fundamental properties of the device itself. Above all, it is important to know if the device obeys the reciprocity theorem, whether it suffers internal power losses, and whether it shows any symmetry relative to its ports. Without such prior knowledge, a complete characterization of the external behaviour of an N-port junction would require in the most general case N^2 complex parameters forming, say, the elements of a matrix. If it is known in advance that the multiport junction in question has the property of reciprocity, suffers no internal losses, and has some symmetry, the number of independent variables can be cut down appreciably.

Prior knowledge about the relations between the matrix coefficient of the multiport junction in question offers an opportunity to avoid errors in setting up ideal matrices that might otherwise occur due to the violation of some physical laws. In the analysis of real matrices, prior knowledge about the property of reciprocity, freedom from internal power losses, and existence of symmetry can be utilized for a check on the results of the computations and on the severity of random errors in measurements.

3.6. Reciprocal Multiport Junctions

Reciprocal multiport junctions are those which obey the reciprocity theorem for any two ports, with any arbitrary conditions existing at all other ports. The reciprocity theorem known from electrodynamics [see Eq. (A.9) in the Appendix] has the following corollary: If an emf at one port of a multiport junction gives rise to an electric current in another, short-circuited port, placing the emf source in the other port will give rise to a similar current in the first port if it is short-circuited, or mathematically

$$\dot{I}_2/\dot{V}_1 = \dot{I}_1/\dot{V}_2$$

If equalities similar to the above one hold for ports m and n, with all the other ports short-circuited, all non-normalized admittance matrix elements arranged symmetrically about the main diagonal will be equal pairwise,

$$\dot{Y}_{mn} = \dot{Y}_{nm}$$

Hence, the non-normalized admittance matrix is symmetrical:

$$\tilde{\mathbf{Y}} = \tilde{\mathbf{Y}}_t$$

The change-over to normalized voltages and currents does not affect the above statement in any way, and a similar reciprocity relation holds for normalized admittance matrices as well:

$$\mathbf{Y} = \mathbf{Y}_t$$

or

$$\dot{y}_{mn} = \dot{y}_{nm} \tag{3.16}$$

And, finally, because the rules for finding normalized voltages and currents reduce any transmission lines to the same mathematical model in the form of an equivalent long line with a characteristic impedance of unity, the reciprocity relation (3.16) proves likewise valid for multiport junctions with arbitrary incoming transmission lines. Thus, *a necessary condition for the reciprocity of a multiport junction is the symmetry of its normalized admittance matrix.*

The impedance matrix and the scattering matrix of a reciprocal multiport junction are likewise symmetrical:

$$\mathbf{Z} = \mathbf{Z}_t$$

and

$$\mathbf{S} = \mathbf{S}_t$$

The Z-matrix must be symmetrical for reciprocal multiport junctions as the inverse of the symmetrical admittance matrix. The symmetry of the scattering matrix for reciprocal multiport junctions is proved, using the conversion relation, Eq. (3.13):

$$\begin{aligned}\mathbf{S} &= (\mathbf{E} - \mathbf{Y})(\mathbf{E} + \mathbf{Y})^{-1} = (\mathbf{E} - \mathbf{Y}_t)(\mathbf{E} + \mathbf{Y}_t)^{-1} \\ &= (\mathbf{E} - \mathbf{Y})_t (\mathbf{E} + \mathbf{Y})_t^{-1} = [(\mathbf{E} + \mathbf{Y})^{-1}(\mathbf{E} - \mathbf{Y})]_t = \mathbf{S}_t\end{aligned}$$

Here, we have used the rule for the transposition of the product of two matrices:

$$[\mathbf{AB}]_t = \mathbf{B}_t \mathbf{A}_t$$

The symmetry displayed by the matrices of a reciprocal multiport junction substantially reduces the number of parameters required to characterize it. For a complete description of a reciprocal N-port junction it will suffice to have as few as $N(N + 1)/2$ complex variables—they are the elements lying on one side of the main diagonal of the matrix, including the main diagonal as well.

It is known from electrodynamics that passive devices owe their property of reciprocity to the fact that they do not contain any anisotropic electromagnetic media, such as biased ferrites or a plasma.

Therefore, in order to ascertain the property of reciprocity, there is no need, in most cases, to undertake a special study; it can be stipulated *a priori*, that is, prior to detailed calculations or measurements on the circuit in question.

3.7. Non-Dissipative Multiport Junctions

The term *non-dissipative* (or *lossless*) refers to multiport junctions free from internal losses of electromagnetic energy. Strictly speaking, absolutely non-dissipative microwave devices are non-existent: any device dissipates (say, converts to heat) some of the microwave energy it receives. In most cases, however, every effort is made to minimize internal losses as much as practicable, and a lossless (or non-dissipative) device is in effect the limiting case of low-loss circuits. The low level of losses should be understood as being vanishingly small as compared with the total power applied to a given N-port junction. Neglecting internal losses, we can simplify many design relations, and such an approach is therefore very useful in both the analysis and synthesis of microwave devices.

To begin with, let us formulate the property of non-dissipativity of a multiport junction in terms of the impedance matrices. The power applied over each incoming line is

$$P_{in\,m} = \operatorname{Re}[\dot{v}_m \dot{i}_m^*] = \frac{1}{2}(\dot{v}_m \dot{i}_m^* + \dot{v}_m^* \dot{i}_m)$$

Adding together the power applied to all the ports of the junction and using matrix notation, we obtain

$$P_{in} = \sum_{m=1}^{N} P_{in\,m} = \frac{1}{2}\sum_{m=1}^{N}(\dot{v}_m \dot{i}_m^* + \dot{v}_m^* \dot{i}_m) = \frac{1}{2}(\langle \mathbf{i}^* \mathbf{v}\rangle + \langle \mathbf{v}^* \mathbf{i}\rangle) \quad (3.17)$$

Here, $\rangle \mathbf{a} = (\dot{a}_1, \dot{a}_2, \ldots, \dot{a}_N)$ is a transposed column matrix, that is, a row matrix. Using the definition of the impedance matrix, (3.8), in Eq. (3.17), recalling the rule for the transposition of the product of matrices

$$(\mathbf{AB})_t = \mathbf{B}_t \mathbf{A}_t$$

and requiring that $P_{in} = 0$ (which must be so in a lossless multiport junction), we obtain

$$P_{in} = \frac{1}{2}(\langle \mathbf{i}^* \mathbf{Z} \mathbf{i}\rangle + \langle \mathbf{i}^* \mathbf{Z}_t^* \mathbf{i}\rangle) = \frac{1}{2}\langle \mathbf{i}^* [\mathbf{Z} + \mathbf{Z}_t^*]\, \mathbf{i}\rangle = 0$$

Input power must be equal to zero, irrespective of the specific form of excitation, $\mathbf{i}\rangle$, applied to a non-dissipative multiport junction. This can be so only if the impedance matrix is such that

$$\mathbf{Z} + \mathbf{Z}^*{}_t = \mathbf{0}$$

or

$$\mathbf{Z} = -\mathbf{Z}_t^*$$

where 0 is a zero matrix of order N.

A similar condition for a multiport to be non-dissipative may be derived also for the admittance matrix:

$$\mathbf{Y} + \mathbf{Y}_t^* = \mathbf{0}$$

or

$$\mathbf{Y} = -\mathbf{Y}_t^*$$

Example 3.2. A non-dissipative two-port junction. When written at length, the condition for a two-port junction to be non-dissipative takes the form

$$\begin{bmatrix} r_{11}+jx_{11} & r_{12}+jx_{12} \\ r_{21}+jx_{21} & r_{22}+jx_{22} \end{bmatrix} = \begin{bmatrix} -r_{11}+jx_{11} & -r_{21}+jx_{21} \\ -r_{12}+jx_{12} & -r_{22}+jx_{22} \end{bmatrix}$$

Hence,

$$r_{11} = r_{22} = 0$$
$$r_{12} = -r_{21}$$
$$x_{12} = x_{21}$$

Thus, the impedance matrix of a non-dissipative two-port junction contains only four independent variables and may be re-cast as follows:

$$\mathbf{Z} = \mathbf{R} + j\mathbf{X} = \begin{bmatrix} 0 & r_{12} \\ -r_{12} & 0 \end{bmatrix} + j \begin{bmatrix} x_{11} & x_{12} \\ x_{12} & x_{22} \end{bmatrix}$$

As is seen, the impedance matrix of a non-dissipative two-port junction (or, more generally, of an N-port junction) has a skew-symmetrical real part and a symmetrical imaginary part. The same holds for the admittance matrix of a non-dissipative two-port junction.

If, in addition to being non-dissipative, a multiport junction is reciprocal, then, due to the symmetry of its Z and Y matrices, the following identities must take place:

$$\mathbf{R} \equiv \mathbf{0}$$

and

$$\mathbf{G} \equiv \mathbf{0}$$

where 0 is a zero matrix of order N. Therefore, the impedance and admittance matrices of a reciprocal and non-dissipative multiport junction are pure imaginaries:

$$\mathbf{Z} = j\mathbf{X}$$

and

$$\mathbf{Y} = j\mathbf{B}$$

For this reason, reciprocal and non-dissipative multiport junctions are often referred to as *reactive*. A reactive multiport only has as few as $N(N+1)/2$ real variables — they are the elements lying on one side of the main diagonal of the **X** or **B** matrix, including the main diagonal itself.

Now let us define the non-dissipativity of a multiport junction in terms of its scattering matrix. Under the wave approach, the power entering (or leaving) a multiport junction over any of the transmission lines may be regarded as the difference between the incident and reflected power waves. For the mth input line,

$$P_{in\,m} = P_{im} - P_{r\,m} = |\dot{v}_{im}|^2 - |\dot{v}_{r\,m}|^2$$

Adding together the power at all the ports of an N-port junction and using matrix notation, we obtain

$$P_{in} = \sum_{m=1}^{N} P_{in\,m} = \sum_{m=1}^{N} (|\dot{v}_{i\,m}|^2 - |\dot{v}_{r\,m}|^2) = \langle \mathbf{v}_i^* \mathbf{v}_i \rangle - \langle \mathbf{v}_r^* \mathbf{v}_r \rangle$$

Using Eq. (3.2a) which defines the scattering matrix and requiring that $P_{in} = 0$ (which must be the case in a non-dissipative multiport junction), we get

$$P_{in} = \langle \mathbf{v}_i^* \mathbf{v}_i \rangle - \langle \mathbf{v}_i^* \mathbf{S}_i^* \mathbf{S} \mathbf{v}_i \rangle = \langle \mathbf{v}_i^* [\mathbf{E} - \mathbf{S}_i^* \mathbf{S}] \mathbf{v}_i \rangle = 0$$

where **E** is an identity (unit) matrix of order N. The left-hand side of the above equation must be equal to zero with any distribution of incident voltage waves. This will happen if the scattering matrix of a non-dissipative multiport junction satisfies the *condition of unitarity*

$$\mathbf{S}_i^* \mathbf{S} = \mathbf{E}$$

In the special case of a one-port, the condition of unitarity reduces to an obvious statement

$$|\dot{\rho}| = 1$$

Unitary matrices show a number of specific properties. The norm of each column in a unitary matrix (that is, the square root of the sum of the squares of the magnitudes of the column elements) is unity, the columns are mutually orthogonal ($\langle \mathbf{S}^{*(m)} \mathbf{S}^{(n)} \rangle = 0$ for $m \neq n$), whereas the determinant of a unitary matrix has a magnitude of unity and may be written as

$$\det \mathbf{S} = \exp(j\varphi)$$

Let us turn to a few examples.

Example 3.3. A non-dissipative two-port junction. When written at length, the condition for a second-order scattering matrix to be

unitary reduces to the following equalities:

$$|\dot{s}_{11}|^2 + |\dot{s}_{21}|^2 = 1$$
$$|\dot{s}_{22}|^2 + |\dot{s}_{12}|^2 = 1$$
$$\dot{s}_{11}^* \dot{s}_{12} + \dot{s}_{21}^* \dot{s}_{22} = 0$$

The first two lines are obvious expressions for the law of conservation of energy when a non-dissipative two-port junction is driven via port *1* and port *2*, with the opposite port terminated in a matched load. The third line establishing a further relation between the amplitudes and phases of the scattering coefficients is less obvious. From solving the three equalities simultaneously it follows that any non-dissipative two-port junction must satisfy the following constraints

$$|\dot{s}_{11}| = |\dot{s}_{22}|$$
$$|\dot{s}_{12}| = |\dot{s}_{21}|$$
$$\varphi_{11} + \varphi_{22} = \varphi_{12} + \varphi_{21} \pm \pi$$

where φ_{mn} is the phase of the (mn)th scattering coefficient.

Thus, for a non-dissipative two-port junction (reciprocal or non-reciprocal), the magnitudes of the transmission coefficients in both directions and the magnitudes of the reflection coefficients at each port are pairwise equal, and the phases of all the scattering coefficients are not independent variables.

The constraints defined above should be taken into account when writing an ideal scattering matrix for a non-dissipative two-port junction. Assuming that the four independent variables are

$$|\dot{s}_{11}| = \cos\tau$$
$$\varphi_{11} = \varphi_1$$
$$\varphi_{12} = \varphi_2$$

and

$$\varphi_{21} = \varphi_3$$

we arrive at the canonical form of the scattering matrix for a non-dissipative two-port junction:

$$\mathbf{S} = \begin{bmatrix} \cos\tau \exp(j\varphi_1) & \sin\tau \exp(j\varphi_2) \\ \sin\tau \exp(j\varphi_3) & -\cos\tau \exp j(\varphi_2 + \varphi_3 - \varphi_1) \end{bmatrix} \tag{3.18}$$

If, in addition to being non-dissipative, a two-port junction is reciprocal, it is required that $\varphi_2 = \varphi_3$. Then the number of independent variables of a reactive (that is, both reciprocal and non-dissipative)

two-port junction is reduced to three:

$$0 \leqslant \tau \leqslant \pi/2$$
$$0 \leqslant \varphi_1 \leqslant 2\pi$$

and

$$0 \leqslant \varphi_2 \leqslant 2\pi$$

Example 3.4. An ideal circulator as a non-dissipative three-port junction matched at all the ports. Suppose that, as shown in Fig. 3.3*a*, the non-dissipative three-port junction has at its ports adjustable reactive matching elements, which may be movable discontinuities

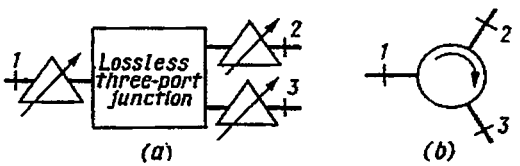


Fig. 3.3. To matching a lossless three-port junction: *a*, block diagram; *b*, graphical symbol for a circulator

of the Tatarinov type. We inquire if it is possible to tune these matching elements so that the ports of the three-port junction can be matched all at the same time. In other words, we inquire if it is possible to realize the scattering matrix of the compound three-port junction (the one which includes the matching elements) in the following form:

$$\mathbf{S} = \begin{bmatrix} 0 & \dot{s}_{12} & \dot{s}_{13} \\ \dot{s}_{21} & 0 & \dot{s}_{23} \\ \dot{s}_{31} & \dot{s}_{32} & 0 \end{bmatrix}$$

where the desired matched condition is indicated by zeros on the main diagonal.

To answer this question, we use the condition for the scattering matrix to be unitary and write the concommitant equalities for the scattering coefficients

$$\left.\begin{aligned} |\dot{s}_{21}|^2 + |\dot{s}_{31}|^2 &= 1 \\ |\dot{s}_{12}|^2 + |\dot{s}_{32}|^2 &= 1 \\ |\dot{s}_{13}|^2 + |\dot{s}_{23}|^2 &= 1 \\ \dot{s}_{31}^* \dot{s}_{32} &= 0 \\ \dot{s}_{21}^* \dot{s}_{23} &= 0 \\ \dot{s}_{13}^* \dot{s}_{12} &= 0 \end{aligned}\right\} \qquad (3.19)$$

Equations (3.19) may be treated as a set of nonlinear equations in the scattering coefficients. This set may have two solutions:

$$\text{(I)}\quad \dot{s}_{31}=0,\ |\dot{s}_{21}|=1,\ \dot{s}_{23}=0,\ |\dot{s}_{13}|=1,\ \dot{s}_{12}=0,\ |\dot{s}_{23}|=1$$

$$\text{(II)}\quad \dot{s}_{32}=0,\ |\dot{s}_{12}|=1,\ \dot{s}_{13}=0,\ |\dot{s}_{23}|=1,\ \dot{s}_{21}=0,\ |\dot{s}_{31}|=1$$

The above solutions correspond to the ideal scattering matrices of matched non-dissipative three-port junctions:

$$\mathbf{S}^{\mathrm{I}}=\begin{bmatrix} 0 & 0 & \exp(j\varphi_3) \\ \exp(j\varphi_1) & 0 & 0 \\ 0 & \exp(j\varphi_2) & 0 \end{bmatrix}$$

$$\mathbf{S}^{\mathrm{II}}=\begin{bmatrix} 0 & \exp(j\varphi_2) & 0 \\ 0 & 0 & \exp(j\varphi_3) \\ \exp(j\varphi_1) & 0 & 0 \end{bmatrix} \qquad (3.20)$$

where the arbitrary phase constants φ_1 through φ_3 do not affect the unitary property of the scattering matrix and are solely dependent on the location of the terminal planes on the three-port junction.

As follows from Eq. (3.20), the microwave power applied to any one port can pass on to only one of the other ports. In case I the sequence in power transfer is from port *1* to port *2* to port *3*. In case II, the sequence in power transfer is the other way around. Devices having ideal scattering matrices defined in (3.20) are called *three-port circulators*. Their graphical symbol is shown in Fig. 3.3*b*. Circulators are widely used to separate the input and output signals of parametric amplifiers and to isolate the outputs of a transmitter and the input of a receiver operating into a common antenna. Microwave circulators are non-reciprocal devices and their construction requires the use of a gyrotropic material (ordinarily, a biased ferrite) inside a three-port junction.

On the basis of Example 3.4, we may formulate the following theorem: *Any non-reciprocal, non-dissipative three-port junction can be converted to an ideal circulator by appropriately tuning the reactive matching elements at its ports.* Several designs of circulators are discussed in Chap. 6.

Analysis of (3.19) would also show that this set of equations has no solution at all in the case of reciprocal circuits with symmetrical scattering matrices, that is, when $\dot{s}_{12}=\dot{s}_{21}$, $\dot{s}_{13}=\dot{s}_{31}$ and $\dot{s}_{23}=\dot{s}_{32}$. This result can be stated as a lemma: *A reciprocal non-dissipative three-port junction cannot be matched at all of its ports at the same time with the aid of reactive matching elements.*

Example 3.5. An ideal directional coupler in the form of a reactive four-port junction match-terminated at all of its ports. Let a reac-

tive (that is, reciprocal and non-dissipative) four-port junction have adjustable reactive matching devices at its ports as shown in Fig. 3.4. We set out to see what form of scattering matrix the resultant composite four-port junction will have, assuming that a perfect

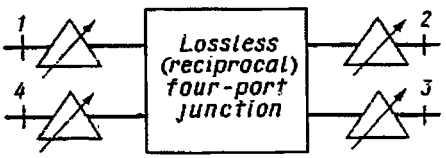


Fig. 3.4. To matching a lossless four-port junction

match can be obtained by suitably tuning the matching elements. In view of the property of reciprocity, the scattering matrix will, after a perfect match has been secured, take up the following form:

$$S=\begin{bmatrix} 0 & \dot{s}_{21} & \dot{s}_{31} & \dot{s}_{41} \\ \dot{s}_{21} & 0 & \dot{s}_{32} & \dot{s}_{42} \\ \dot{s}_{31} & \dot{s}_{32} & 0 & \dot{s}_{43} \\ \dot{s}_{41} & \dot{s}_{42} & \dot{s}_{43} & 0 \end{bmatrix}$$

Since this is a unitary matrix, the following equations may be written:

$$\left.\begin{aligned} |\dot{s}_{21}|^2+|\dot{s}_{31}|^2+|\dot{s}_{41}|^2=1 \\ |\dot{s}_{21}|^2+|\dot{s}_{32}|^2+|\dot{s}_{42}|^2=1 \\ |\dot{s}_{31}|^2+|\dot{s}_{32}|^2+|\dot{s}_{43}|^2=1 \\ |\dot{s}_{41}|^2+|\dot{s}_{42}|^2+|\dot{s}_{43}|^2=1 \end{aligned}\right\} \tag{3.21}$$

$$\dot{s}^*_{31}\dot{s}_{32}+\dot{s}^*_{41}\dot{s}_{42}=0 \tag{3.22}$$

$$\dot{s}^*_{21}\dot{s}_{32}+\dot{s}^*_{41}\dot{s}_{43}=0 \tag{3.23}$$

$$\dot{s}^*_{21}\dot{s}_{42}+\dot{s}^*_{31}\dot{s}_{43}=0 \tag{3.24}$$

On the strength of (3.21), at least one element in each column of the scattering matrix is non-zero. Equations (3.23) and (3.24) in the scattering coefficients $\dot{s}^*_{21}$ and $\dot{s}_{43}$ have an only trivial solution

$$\left.\begin{aligned} \dot{s}_{21}=0 \\ \dot{s}_{43}=0 \end{aligned}\right\} \tag{3.25}$$

if the determinant of this set of equations

$$\dot{\Delta}_1 = \dot{s}_{32}\dot{s}_{31}^* - \dot{s}_{42}\dot{s}_{41}^*$$

is non-zero. If, on the other hand, the determinant $\dot{\Delta}_1$ is zero, a multitude of non-zero solutions exist for the scattering coefficients $\dot{s}_{21}$ and $\dot{s}_{43}$, but another pair of scattering coefficients must be zero because, in view of Eq. (3.22),

$$\dot{s}_{31}^*\dot{s}_{32}^* = \dot{s}_{42}\dot{s}_{41} = 0 \tag{3.26}$$

In such a case, the solutions of Eqs. (3.22) and (3.23) or of (3.22) and (3.24), respectively, are

$$\dot{s}_{32} = 0, \quad \dot{s}_{41} = 0 \tag{3.27}$$

$$\dot{s}_{42} = 0, \quad \dot{s}_{31} = 0 \tag{3.28}$$

on the condition that the determinant

$$\dot{\Delta}_2 = \dot{s}_{31}^*\dot{s}_{43} - \dot{s}_{42}\dot{s}_{21}^*$$

or

$$\dot{\Delta}_3 = \dot{s}_{41}^*\dot{s}_{43} - \dot{s}_{32}\dot{s}_{21}^*$$

is non-zero. Note that, subject to Eqs. (3.22) through (3.24), the determinants $\dot{\Delta}_1$, $\dot{\Delta}_2$ and $\dot{\Delta}_3$ may turn to zero only if Eq. (3.27) or (3.28), (3.25) or (3.28), (3.25) or (3.27) is satisfied. Thus, the assumption about an ideal match at the ports of a reactive four-port junction has as its outcome the fact that its scattering matrix should have one of the three forms

$$\begin{bmatrix} 0 & 0 & \dot{s}_{31} & \dot{s}_{41} \\ 0 & 0 & \dot{s}_{32} & \dot{s}_{42} \\ \dot{s}_{31} & \dot{s}_{32} & 0 & 0 \\ \dot{s}_{41} & \dot{s}_{42} & 0 & 0 \end{bmatrix}$$

or

$$\begin{bmatrix} 0 & \dot{s}_{21} & \dot{s}_{31} & 0 \\ \dot{s}_{21} & 0 & 0 & \dot{s}_{42} \\ \dot{s}_{31} & 0 & 0 & \dot{s}_{43} \\ 0 & \dot{s}_{42} & \dot{s}_{43} & 0 \end{bmatrix}$$

or

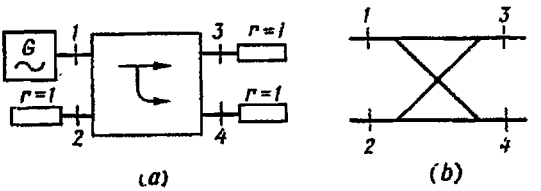
$$\begin{bmatrix} 0 & \dot{s}_{21} & 0 & \dot{s}_{41} \\ \dot{s}_{21} & 0 & \dot{s}_{32} & 0 \\ 0 & \dot{s}_{32} & 0 & \dot{s}_{43} \\ \dot{s}_{41} & 0 & \dot{s}_{43} & 0 \end{bmatrix}$$

In any of them, there are two zeros in each column of the scattering matrix. On changing the numbering scheme for the ports (see Eq. (3.14)), the three forms may be reduced to a single common form

$$\mathbf{S} = \left[\begin{array}{cc:cc} 0 & 0 & \dot{s}_{31} & \dot{s}_{41} \\ 0 & 0 & \dot{s}_{32} & \dot{s}_{42} \\ \hdashline \dot{s}_{31} & \dot{s}_{32} & 0 & 0 \\ \dot{s}_{41} & \dot{s}_{42} & 0 & 0 \end{array}\right] = \left[\begin{array}{c:c} \mathbf{0} & \mathbf{T}_t \\ \hdashline \mathbf{T} & \mathbf{0} \end{array}\right] \tag{3.29}$$

Here, the scattering matrix S has been decomposed into four submatrices; in this way, we have changed to a *blocked scattering matrix.*

The four-port junction answering the scattering matrix defined by Eq. (3.29) can conveniently be pictured as a feed-through device



(a) (b)

Fig. 3.5. Directional coupler in a microwave transmission system: *a*, block diagram; *b*, graphical symbol

(Fig. 3.5) with two pairs of ports, namely left-hand ports *1* and *2*, and right-hand ports *3* and *4*. Each pair of ports is matched and decoupled (isolated) from the other. A microwave four-port waveguide junction in which the ports are pairwise decoupled and matched and which is described by an ideal scattering matrix of the form defined by (3.29) is called a *directional coupler*. The off-diagonal block T in the scattering matrix of a directional coupler contains the wave transmission coefficients from one pair of ports to the other, and it may therefore be called a *transmission* (or *transfer*) *block*. The transmission block must be unitary, which is easy to ascertain by apply-

ing the condition of unitarity to the scattering matrix in (3.29):

$$\begin{bmatrix} 0 & \mathbf{T}_t^* \\ \mathbf{T}^* & 0 \end{bmatrix} \begin{bmatrix} 0 & \mathbf{T}_t \\ \mathbf{T} & 0 \end{bmatrix} = \begin{bmatrix} \mathbf{E} & 0 \\ 0 & \mathbf{E} \end{bmatrix}$$

Hence,

$$\mathbf{T}_t^* \mathbf{T} = \mathbf{E}$$

The general form of the transmission block is in agreement with (3.18). Through a proper shift in the location of the terminal planes at the ports of a directional coupler (see Eq. (3.15)), it is possible to reduce the transmission block to one of two forms

$$\mathbf{T}^{\mathrm{I}} = \begin{bmatrix} \cos\tau & \sin\tau \\ \sin\tau & -\cos\tau \end{bmatrix}$$

or

$$\mathbf{T}^{\mathrm{II}} = \begin{bmatrix} \cos\tau & j\sin\tau \\ j\sin\tau & \cos\tau \end{bmatrix}$$

where only one independent variable is left: $0 \leqslant \tau \leqslant \pi/2$.

Directional couplers are widely used as building blocks for ramified microwave transmission systems. A valuable property of directional couplers is the fact that all of its ports are matched — a condition unattainable in reactive three-port junctions (see Example 3.4). Proceeding from Example 3.5, we are in a position to state the following important theorem: *Any reactive four-port junction can be converted to an ideal directional coupler with a scattering matrix of the form defined in (3.29) by appropriately tuning the reactive matching devices at its ports.*

3.8. Symmetrical Multiport Junctions

A symmetrical multiport is that for which a change in the numbering scheme for the ports will not affect its parameter matrices. There may be a *geometrical* (*physical*) *symmetry* and an *electrical symmetry*. A purely electrical symmetry (which is not a consequence of a physical symmetry) can be secured through a proper choice of element values for the multiport junction in question and cannot be established *a priori*. In contrast, a physical symmetry can be established in advance (before an electrical calculation is made), and it always brings with it an electrical symmetry. In the case of a physical symmetry, any spatial transformation, technically known as a *covering operation*, will leave a multiport junction similar to itself. Covering operations may be elementary and complex. The former include rotations of a multiport junction about an axis of symmetry and reflection with respect to a plane of symmetry. Covering operations are each a series of consecutive elementary covering operations.

Formally, the symmetry of a multiport junction may be described by a square symmetry operator matrix G having the same order as there are ports. Symmetry operator matrices may be specified by writing

$$G|v_r\rangle = SG|v_i\rangle \tag{3.30}$$

It implies that an arbitrary incident-wave column matrix may be symmetrically transformed in such a way,

$$|v_i\rangle \rightarrow G|v_i\rangle$$

that the respective reflected-wave column matrix will be transformed in exactly the same way,

$$|v_r\rangle \rightarrow G|v_r\rangle$$

Owing to symmetry, the waves can be interchanged only at symmetrical ports, although sometimes this will be accompanied by a change in the original conventional direction of voltage at some ports. Because of this and similarly to the re-numbering matrix introduced in Sec. 3.4, each row and each column of the G matrix should contain a non-zero element which, however, may assume value $+1$ or -1 (when, as a result of a covering operation, the positive direction of voltage at the respective port is reversed). The symmetry matrix must be orthogonal, which implies that it must satisfy the condition

$$G_t G = E$$

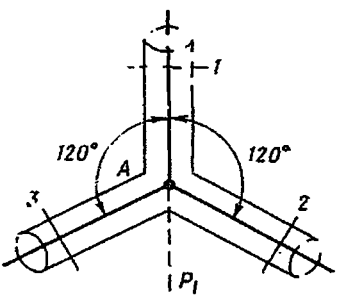


Fig. 3.6. Symmetrical coaxial-line Y-junction

Example 3.6. A symmetrical Y-junction of coaxial waveguides. If the terminal planes are located at equal distances from junction point A (Fig. 3.6), the Y-junction will have both a rotational symmetry and a reflection symmetry. Let us, first, consider the rotational symmetry. Obviously, rotating through 120° clockwise about point A will cause the arms of the junction to interchange positions, which is equivalent to a change in the numbering scheme as follows: port *1* becomes port *2*, port *2* becomes port *3*, and port *3* becomes port *1*. Therefore, the G_1 matrix is identical to the respective renumbering matrix set up by the rules set forth in Sec. 3.4:

$$G_1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

Another symmetry operator matrix can be written on rotating the Y-junction about point A through 120° counter-clockwise, which is

equivalent to a change in the numbering scheme as follows: port *1* becomes port *3*, port *2* becomes port *1*, and port *3* becomes port *2*. Therefore,

$$G_2 = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

Let us now consider a reflection operation on the Y-junction. A reflection with respect to the plane of symmetry P_1 passing through port *1* causes the arms of the junction to interchange positions, while the junction is left intact, but the ports are now numbered differently: port *1* remains port *1*, port *2* becomes port *3*, and port *3* becomes port *2*. The resultant symmetry operator matrix takes the form

$$G_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

Proceeding in a similar way, but effecting reflections with respect to the planes of symmetry passing through port *2* and port *3*, we obtain two more symmetry operator matrices:

$$G_4 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

$$G_5 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

A simple analysis will show that of all the symmetry operator matrices G_1 through G_5, only two are mutually independent (one for each form of symmetry), say, G_1 and G_3. The remaining matrices can be deduced as their products:

$$G_2 = G_1G_1$$
$$G_4 = G_1G_3$$
$$G_5 = G_3G_1$$

All the symmetry operator matrices of a multiport junction (including its unit matrix E) form a group. In mathematics, the term '*group*' is given to a set of objects, including a binary operation defined on this set. In our case, the objects are square matrices, and the binary operation is the usual multiplication of two matrices, which is not commutative. Every group must contain a unit element (in our case, a unit matrix) and every element should have

its inverse (in our case, $\mathbf{G}^{-1} = \mathbf{G}_t$ by virtue of the orthogonality of symmetry operator matrices).

The mutually independent symmetry operator matrices which, when multiplied together, yield all the other symmetry operator matrices of a group are called the *generators* of the group. To completely specify the symmetry properties of a multiport junction, it will suffice to have only the generators of a group. For a symmetrical Y-junction, the generators of the group may be the matrices $\mathbf{G}_1$ and $\mathbf{G}_3$.

Now we consider the case where a need arises to use negative elements in a symmetry operator matrix.

Example 3.7. A series T-junction of two-wire transmission lines (Fig. 3.7). If the terminal planes at ports *1* and *2* are chosen to lie an equal distance from plane P_3 passing through the axis of arm (port) *3*, then ports *1* and *2* are mutually symmetrical. If, however, we mentally replace the right-hand half of the three-port junction by its reflection (because of which the respective electromagnetic fields will likewise 'be covered', that is, interchange their positions in a mirror fashion), the voltage at port *3* will undergo a phase reversal. If we reverse the positive direction of voltage at port *3*, the conditions in the three-port junction and its description will remain unchanged. This means that port *3* is symmetrical to itself, but only if the opposite direction of voltage is taken. Therefore, the symmetry operator matrix for a series T-junction requires a change in the numbering scheme: port *1* becomes port *2*, port *2* becomes port *1*, and port *3* becomes port *3*, which gives

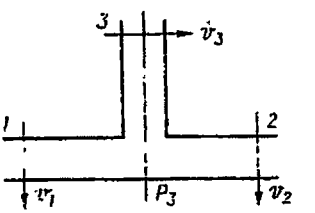


Fig. 3.7. Series T-junction of two-wire transmission lines

$$\mathbf{G} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

This is the only symmetry operator matrix that a series *T*-junction can have.

The use of symmetry operator matrices in the analysis of various microwave devices is based on the fact that they commute with parameter matrices. In other words, the following identities exist:

$$\mathbf{GS} = \mathbf{SG}, \quad \mathbf{GZ} = \mathbf{ZG}, \quad \mathbf{GY} = \mathbf{YG} \tag{3.31}$$

To prove the property defined by (3.31) for the scattering matrix, both sides of (3.30) should be premultiplied by the $\mathbf{G}^{-1}$ matrix. The resultant equality will exist for any distribution of the excitation $\mathbf{v}_t$). Therefore, proceeding from the definition of the scattering

matrix in (3.2a), we may write

$$\mathbf{G}^{-1}\mathbf{S}\mathbf{G} \equiv \mathbf{S}$$

and this leads to the first identity in (3.31). The remaining identities in (3.31) can be proved in a similar way for the impedance and admittance matrices as well.

If the symmetry operator matrices are known or specified in advance, the identities in (3.31) establish relationships between the elements of one and the same parameter matrix of a multiport junction, thereby reducing the number of independent variables that is needed to specify a symmetrical multiport junction.

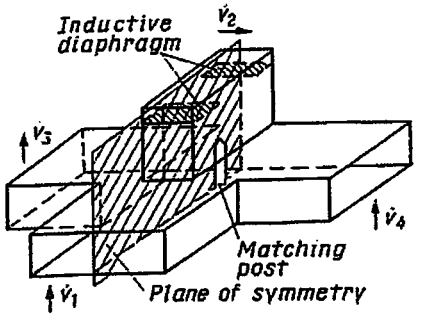


Fig. 3.8. Magic tee

Example 3.8. A hybrid or magic tee (the concurrent use of reciprocity, non-dissipativity and symmetry). The reciprocal multiport junction shown in Fig. 3.8 is a combination of an *H*-plane tee (arm *1*) and an *E*-plane tee (arm *2*), each made in rectangular waveguide operating in the TE_{10} mode. If the dielectric used to fill the waveguides is isotropic, the device will have one plane of symmetry which cuts arms *1* and *2*, and its symmetry operator and scattering matrices will have the form

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

$$\mathbf{S} = \begin{bmatrix} \cdot & \cdot & \cdot & \cdot \\ S_{11} & S_{12} & S_{13} & S_{14} \\ \cdot & \cdot & \cdot & \cdot \\ S_{12} & S_{22} & S_{23} & S_{24} \\ \cdot & \cdot & \cdot & \cdot \\ S_{13} & S_{23} & S_{33} & S_{34} \\ \cdot & \cdot & \cdot & \cdot \\ S_{14} & S_{24} & S_{34} & S_{44} \end{bmatrix}$$

In writing the symmetry operator matrix, it is taken into account that ports *3* and *4* are mutually symmetrical and ports *1* and *2* are each symmetrical to itself. Also, seen looking into port *2*, this four-port tee is analogous to a series *T*-junction of transmission lines—this fact is taken care of by writing —1 in the second row of the G-matrix (see Example 3.7).

On taking the product of the matrices on the right- and left-hand sides of the identity GS ≡ SG and equating the elements of the products *pq* (the product of the *p*th row in the multiplicand by the *q*th column in the multiplier), we obtain the following relations between the scattering coefficients for a magic tee:

$$\left.\begin{aligned} \dot{s}_{12} &\equiv -\dot{s}_{12} && (p=1,\ q=2)\\ \dot{s}_{14} &\equiv \dot{s}_{13} && (p=3,\ q=1)\\ \dot{s}_{24} &\equiv -\dot{s}_{23} && (p=3,\ q=2)\\ \dot{s}_{33} &\equiv \dot{s}_{44} && (p=4,\ q=3) \end{aligned}\right\} \tag{3.32}$$

The first line in (3.32) is most important of all, as it leads to $\dot{s}_{12} = 0$.

Thus, it follows from the reciprocity and reflection symmetry of a magic-tee that ports *1* and *2* are ideally decoupled from each other at any frequency. A 'bare' magic-tee (the one free from any other elements), when excited at ports *1* and *2*, behaves as a pair of uncoupled *T*-junctions and shows a considerable amount of mismatch between the ports. Proper match at port *1* is usually obtained with a tuning screw placed in the plane of symmetry (see Fig. 3.8), proper match at port *2* is obtained with an inductive diaphragm. Because ports *1* and *2* are decoupled, the two tuning elements act independently of each other. Therefore, through a proper choice of their locations and dimensions, ports *1* and *2* can be completely matched (of course, provided ports *3* and *4* are match-terminated).

Assuming that ports *1* and *2* are matched and the conditions defined by Eq. (3.32) are satisfied, the scattering matrix of a magic-tee takes the form

$$\mathbf{S} = \begin{bmatrix} 0 & \mathbf{T}_t \\ \mathbf{T} & \mathbf{R} \end{bmatrix}$$

where

$$\mathbf{T} = \begin{bmatrix} \dot{s}_{13} & \dot{s}_{23} \\ \dot{s}_{13} & -\dot{s}_{23} \end{bmatrix}$$

and

$$\mathbf{R} = \begin{bmatrix} \dot{s}_{33} & \dot{s}_{34} \\ \dot{s}_{34} & \dot{s}_{33} \end{bmatrix}$$

In the absence of losses, the above scattering matrix must be unitary, which fact leads to the following equalities:

$$\mathbf{T}_t^* \mathbf{T} = \mathbf{E}$$

and

$$\mathbf{T}_t^* \mathbf{R} = 0$$

Premultiplying the second of the above equalities by the T matrix and taking into account the first equality, we obtain

$$\mathbf{R} \equiv 0$$

that is,

$$\dot{s}_{33} = \dot{s}_{34} = 0$$

It follows then that ports *3* and *4* are likewise mutually decoupled and matched. This result is of a general character and may be stated as a theorem: *If a reactive four-port junction is matched and at the same time decoupled at one pair of ports, it will of necessity be matched and decoupled at the other pair of ports.* In other words, such a four-port is a *directional coupler.*

With a proper choice of locations for the terminal planes, the scattering matrix for a matched magic tee takes the form

$$\mathbf{S} = \begin{bmatrix} 0 & \mathbf{T}_t \\ \mathbf{T} & 0 \end{bmatrix}$$

where

$$\mathbf{T} = \frac{\exp(j\varphi)}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

Thus, a matched magic tee is an equal power divider, that is, a directional coupler which divides input power into two equal parts between any pair of outgoing transmission lines such that at one decoupled port the split power is in phase and at the other it is in anti-phase.

Equal power dividers are a class of 3-dB directional couplers which are sometimes called bridge junctions. It would appear reasonable to call the hybrid tee described above as a *double T-junction.* However, the usual name is the *hybrid junction* or the *bridge hybrid.* When accurately tuned, a magic tee can have a relative bandwidth of 10 to 15% of the centre operating frequency.

Chapter Four

Composite Microwave Multiport Junctions

4.1. The Decomposition Principle in the Analysis of Microwave Multiport Junctions

The most versatile technique used to calculate the electrical performance of microwave devices is the decomposition principle. With it, a complex device is decomposed into a number of simpler circuits amenable to an independent analysis and definable in terms of specific parameter matrices. We shall call such simpler circuits

the *basic elements.* If the basic elements have been analysed and the values of the quantities defining the parameter matrix of each basic element have been determined by a previous calculation, the analysis of the microwave system made up of these basic elements reduces to evaluating the parameter matrices by a suitable algorithm developed to combine two or more basic elements.

For low-frequency circuits it would suffice to use a set of basic elements which include a resistor (a power absorber), a capacitor (an electric energy storage device), and an inductor (a magnetic energy storage device). At microwave frequencies, electromagnetic energy can be stored and absorbed by any element of the entire volume taken up by the complex microwave device, and it is not at all easy to identify the basic elements.

The traditional approach to the decomposition of microwave networks consists in that each selected basic element is replaced by an equivalent circuit made up of lumped-parameter elements L, C and R, and of transmission-line sections. The necessary electrodynamic calculations of the basic elements are made in advance, and the results are presented as approximate equations and reference tables which relate the element values adopted in the equivalent circuits to the physical dimensions of the basic element involved, wavelength, and the parameters of the ferrodielectrics used. Among the advantages of this approach are versatility, close analogy with the theory of low-frequency circuits, and an easy-to-grasp presentation of what a given composite microwave circuit can do and how, all of this achievable through a reasonable idealization of equivalent circuits. On the demerit side of the traditional approach are the loss of accuracy due to the use of simplified equivalent circuits, and certain difficulties in evaluating the magnitude of error in the calculations.

These disadvantages are readily circumvented with a formal electrodynamic approach oriented towards the use of large computers. With this approach, a composite microwave network is decomposed into a number of basic elements in the form of geometrical configurations permitting the parameter matrices to be evaluated either analytically or numerically, solving Maxwell's equations under the specified boundary conditions. Subsequently, the parameter matrix of the composite microwave network is evaluated by the same combining algorithms as are used in the traditional approach based on equivalent circuits. In principle, the electrodynamic approach permits calculations to be carried out to any degree of accuracy, but the analysis itself is less tangible, and each computational program can encompass a narrower class of circuits.

Since no fundamental differences exist between the traditional and the electrodynamic approaches, the subsequent exposition of microwave circuit analysis by decomposition will rely on the traditional approach involving equivalent circuits for the basic elements.

Two levels of decomposition will be used, namely:

(1) The basic elements of a given composite microwave network will be presented as equivalent circuits made up of transmission-line sections and *LCR* elements.

(2) An entire microwave transmission system will be reduced to a combination of simple multiport junctions and analyzed, using multiport-combining algorithms.

4.2. Analysis of Cascaded Two- and One-Port Junctions by Use of Cascading Matrices

Many microwave two- and one-port junctions are cascaded connections. In a cascaded connection, the output of a preceding two-port is the input for the next, as shown in Fig. 4.1. Analysis of

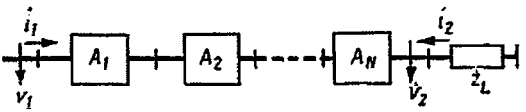


Fig. 4.1. Cascade connection of two-ports

a cascaded connection can markedly be simplified if we describe the individual two-ports and the entire composite junction in terms of what are known as *cascading (or transformation) matrices.*

In setting up a cascading matrix, the excitation applied to the two-port in question is chosen to be a pair of electric quantities defining the condition at one port (as a rule, this is the second port), and the response of the two-port is a corresponding pair of electrical quantities defining the condition at the other port (this usually is the first port).

If we take the *classical cascading matrix*, which we will call the A matrix*, the excitation and response of a two-port are connected by a relation of the form

$$\begin{bmatrix} \dot{v}_1 \\ \dot{i}_1 \end{bmatrix} = \begin{bmatrix} \dot{a} & \dot{b} \\ \dot{c} & \dot{d} \end{bmatrix} \begin{bmatrix} \dot{v}_2 \\ -\dot{i}_2 \end{bmatrix}, \quad \mathbf{A} \equiv \begin{bmatrix} \dot{a} & \dot{b} \\ \dot{c} & \dot{d} \end{bmatrix} \tag{4.1}$$

With this definition, the cascading matrix of any number of cascaded two-ports is equal to the product of the cascading matrices of the individual two-ports. This is the key property of any cascading matrix. The cascading matrices of cascaded two-ports should be multiplied in the same order as the individual two-ports are cascaded.

* More commonly, it is called the ABCD matrix.—*Translator's note.*

Sometimes, preference is given to what is known as the *wave transformation matrix* T defined as

$$\begin{bmatrix} \dot{v}_{i1} \\ \dot{v}_{r1} \end{bmatrix} = \begin{bmatrix} \dot{t}_{11} & \dot{t}_{12} \\ \dot{t}_{21} & \dot{t}_{22} \end{bmatrix} \begin{bmatrix} \dot{v}_{r2} \\ \dot{v}_{i2} \end{bmatrix}$$

$$\mathbf{T} \equiv \begin{bmatrix} \dot{t}_{11} & \dot{t}_{12} \\ \dot{t}_{21} & \dot{t}_{22} \end{bmatrix}$$

The T matrix does not have any additional advantages over the A matrix.

It follows from (4.1) that the coefficients $\dot{a}$ and $\dot{c}$ of the A matrix are, respectively, the normalized voltage transfer ratio (or normalized voltage gain) and the normalized mutual admittance with the output port, *2*, open-circuited. The coefficients $\dot{b}$ and $\dot{d}$ are, respectively, the normalized mutual impedance and the normalized current transfer ratio (or normalized current gain), with the output port, *2*, short-circuited.

Knowing the coefficients of the A matrix, it is an easy matter to analyse the cascade the last stage of which is loaded in a device with a normalized impedance equal to

$$\dot{z}_L = \dot{v}_2/(-\dot{i}_2)$$

The input impedance of a cascaded one-port like that is, according to (4.1), given by

$$\dot{z}_{in} = \dot{v}_1/\dot{i}_1 = \frac{\dot{a}\dot{v}_2 - \dot{b}\dot{i}_2}{\dot{c}\dot{v}_2 - \dot{d}\dot{i}_2} = \frac{\dot{a}\dot{z}_L + \dot{b}}{\dot{c}\dot{z}_L + \dot{d}} \tag{4.2}$$

where use is made of the elements of the matrix $\mathbf{A} = \mathbf{A}_1\mathbf{A}_2 \ldots \mathbf{A}_N$.

If the termination is a short- or an open-circuit, Eq. (4.2) takes a simpler form

$$\dot{z}_{sc} = \dot{b}/\dot{d} \text{ for } \dot{z}_L = 0 \tag{4.3}$$

or

$$\dot{z}_{oc} = \dot{a}/\dot{c} \text{ for } \dot{z}_L \text{ tending to infinity}$$

The reflection coefficient of a cascaded one-port junction can be found by the usual equation

$$\dot{\rho}_{in} = (\dot{z}_{in} - 1)/(\dot{z}_{in} - 1)$$

on substituting for $\dot{z}_{in}$ from either (4.2) or (4.3).

Conversion between the A matrix and the scattering matrix of two-port junction. To derive equations for conversion from the

cascading matrix to the scattering matrix S, the column vectors representing the excitation and the response in (4.1) should be expressed in terms of the incident and reflected waves

$$\dot{v}_{1,2} = \dot{v}_{i1,2} + \dot{v}_{r1,2}$$

and

$$\dot{i} = \dot{v}_{i1,2} - \dot{v}_{r1,2}$$

On collecting the like terms, we get

$$\left.\begin{aligned} \dot{v}_{i1} + \dot{v}_{r1} &= (\dot{a} - \dot{b})\,\dot{v}_{i2} + (\dot{a} + \dot{b})\,\dot{v}_{r2} \\ \dot{v}_{i1} - \dot{v}_{r1} &= (\dot{c} - \dot{d})\,\dot{v}_{i2} + (\dot{c} + \dot{d})\,\dot{v}_{r2} \end{aligned}\right\} \tag{4.4}$$

Then using Eq. (3.3) which defines the scattering coefficients, and combining Eqs. (4.4) pairwise, we obtain the following conversion equations:

$$\begin{aligned} \dot{s}_{11} &= \frac{(\dot{a} - \dot{d}) + (\dot{b} - \dot{c})}{\dot{a} + \dot{b} + \dot{c} + \dot{d}} \\ \dot{s}_{12} &= \frac{2\,(\dot{a}\dot{d} - \dot{b}\dot{c})}{\dot{a} + \dot{b} + \dot{c} + \dot{d}} \\ \dot{s}_{21} &= \frac{2}{\dot{a} + \dot{b} + \dot{c} + \dot{d}} \\ \dot{s}_{22} &= \frac{-(\dot{a} - \dot{d}) + (\dot{b} - \dot{c})}{\dot{a} + \dot{b} + \dot{c} + \dot{d}} \end{aligned} \tag{4.5}$$

Elementary reciprocal two-port junctions. In order to decompose most reciprocal microwave two-port junctions set up in a cascaded structure it will suffice to have four basic elements which we will also call *elementary two-port junctions*. Their equivalent circuits and classical cascading matrices are given in Fig. 4.2. Note that each elementary two-port junction can be specified by giving only one design quantity (complex or real).

Let us see how the classical cascading matrices are set up for the elementary two-port junctions. The first two-port junction (Fig. 4.2*a*) is a *uniform transmission-line section* of length l. The classical cascading coefficients $\dot{a}$ and $\dot{d}$ of this two-port junction define the voltage distribution when the line is open-circuited and the current distribution when the line is short-circuited. Therefore, they are the same and, if the line is free from losses, equal to $\cos \beta l$. The cascading coefficients $\dot{b}$ and $\dot{c}$ for a lossless transmission-line section are found

as follows:

$$\dot{b} = \left(\frac{\dot{v}_1}{-\dot{i}_2}\right)_{v_2=0} = \left(\frac{\dot{v}_1}{\dot{i}_1}\right)\left(\frac{\dot{i}_1}{-\dot{i}_2}\right)_{v_2=0} = \dot{z}_{sc}\dot{d} = j\tan\beta l\cos\beta l = j\sin\beta l$$

$$\dot{c} = \left(\frac{\dot{i}_1}{\dot{v}_2}\right)_{i_2=0} = \left(\frac{\dot{i}_1}{\dot{v}_1}\right)\left(\frac{\dot{v}_1}{\dot{v}_2}\right)_{i_2=0} = \dot{y}_{oc}\dot{a} = j\tan\beta l\cos\beta l = j\sin\beta l$$

where

$$\dot{z}_{sc} = j\tan\beta l = \dot{y}_{oc}$$

designates the input impedance and the input admittance of the line when its output is short-circuited and open-circuited, respectively.

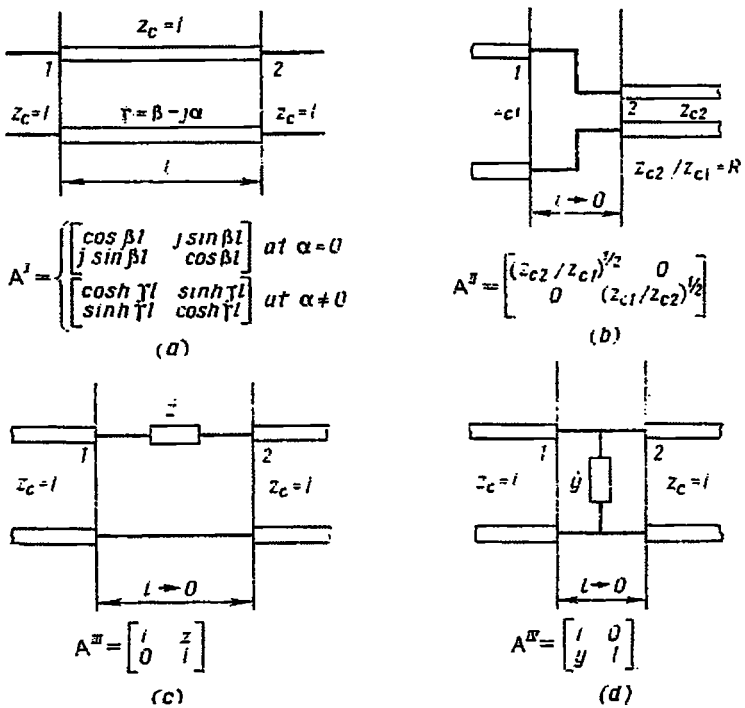

$z_c = l$
1
2
$\dot{\gamma} = \beta - j\alpha$
l
$A^I = \begin{cases} \begin{bmatrix} \cos\beta l & j\sin\beta l \\ j\sin\beta l & \cos\beta l \end{bmatrix} & at\ \alpha = 0 \\ \begin{bmatrix} \cosh\dot{\gamma}l & \sinh\dot{\gamma}l \\ \sinh\dot{\gamma}l & \cosh\dot{\gamma}l \end{bmatrix} & at\ \alpha \neq 0 \end{cases}$
(a)
z_{c2}
$z_{c2}/z_{c1} = R$
$l \to 0$
$A^{II} = \begin{bmatrix} (z_{c2}/z_{c1})^{1/2} & 0 \\ 0 & (z_{c1}/z_{c2})^{1/2} \end{bmatrix}$
(b)
$\dot{z}$
$A^{III} = \begin{bmatrix} 1 & \dot{z} \\ 0 & 1 \end{bmatrix}$
(c)
$\dot{y}$
$A^{IV} = \begin{bmatrix} 1 & 0 \\ \dot{y} & 1 \end{bmatrix}$
(d)

Fig. 4.2. Elementary reciprocal two-ports

For a lossy line section ($\alpha \neq 0$), the cascading coefficients are found in a similar way.

In an elementary two-port junction in the form of a *butt joint* between *two transmission lines* differing in characteristic impedances, z_{c1}

and z_{c2} (Fig. 4.2*b*), the total non-normalized voltages and currents are equal in the plane of the device, that is,

$$\dot{V}_1 = \dot{V}_2$$

and

$$\dot{I}_1 = -\dot{I}_2$$

The '—' sign indicates that the currents enter the junction at each port. Converting the voltages and currents in the above equalities to their normalized values by resort to Eq. (1.15), we obtain a set of equations

$$\dot{v}_1 = \dot{v}_2 (z_{c2}/z_{c1})^{1/2}$$

and

$$\dot{i}_1 = -\dot{i}_2 (\dot{z}_{c1}/z_{c2})^{1/2}$$

from which we readily set up the classical cascading (or ABCD) matrix for the junction.

In an elementary two-port junction in the form of a *lumped series impedance* $\dot{z}$ interposed between two identical transmission lines (Fig. 4.2*c*), the normalized voltage at port *1* is, as follows from Ohm's law, equal to

$$\dot{v}_1 = \dot{v}_2 - \dot{z}\dot{i}_2$$

Also, $\dot{i}_1 = -\dot{i}_2$. These two equalities define the respective cascading coefficients, that is, the elements of the A matrix.

For a two-port junction in the form of a *lumped shunt admittance* $\dot{y}$ placed across a uniform transmission line (Fig. 4.2*d*), the following equalities are true:

$$\dot{v}_1 = \dot{v}_2$$

and

$$\dot{i}_1 = \dot{y}\dot{v}_2 - \dot{i}_2$$

These two equalities define the coefficients of the A matrix of the junction.

Note that the elementary two-port junctions shown in Fig. 4.2, except the transmission-line section, have a zero electrical length and are therefore maximally simplified mathematical models. The inevitable propagation delay in real components of a microwave transmission system analysed by resort to such equivalent circuits can readily be accounted for by cascading transmission-line sections at the input and output of each element.

Conditions of reactivity for a two-port junction. As follows from Eq. (4.5) defining conversion between the cascading and scattering matrices, the reciprocity condition $\dot{s}_{12} = \dot{s}_{21}$ leads to the following

equality:

$$\dot{a}\dot{d} - \dot{b}\dot{c} = \det \mathbf{A} = 1$$

Thus, *for reciprocal two-port junctions, the determinant of the cascading matrix must be equal to unity.* Also, from the assumption of no losses in a reciprocal two-port junction it follows that the cascading coefficients $\dot{a}$ and $\dot{d}$ must be *purely real*, and the cascading coefficients $\dot{b}$ and $\dot{c}$ must be *pure imaginaries.* This property can readily be checked by reference to the elementary two-port junction shown in Fig. 4.2 and extended to a cascade of any number of such two-port junctions.

For non-reciprocal two-port junctions, the condition of non-dissipativity (losslessness) is more elaborate and requires that the cascading matrix should take the form

$$\mathbf{A} = \exp(j\varphi) \begin{bmatrix} \alpha & j\beta \\ j\gamma & \delta \end{bmatrix}$$

where

$$\alpha\delta + \beta\gamma = 1$$

and α, β, γ, δ and φ are real numbers.

Conditions for the symmetry and antimetry of two-port junctions. For symmetric two-port junctions, the following equalities should hold:

$$\dot{s}_{11} = \dot{s}_{22}$$

and

$$\dot{s}_{12} = \dot{s}_{21}$$

Using the conversion equation (4.5), it is an easy matter to establish that symmetry exists when the conditions for the cascading coefficients are satisfied, that is,

$$\dot{a} = \dot{d}$$

and

$$\det \mathbf{A} = 1$$

The inverse of a symmetrical two-port junction is what is called an *antimetric two-port junction.* The term 'antimetric' refers to reciprocal two-port junctions for which the input reflection coefficients are equal in magnitude and opposite in phase at any frequency:

$$\dot{s}_{11} = -\dot{s}_{22}$$

and

$$\dot{s}_{12} = \dot{s}_{21}$$

As follows from the conversion equations (4.5), the condition of antimetry, as written in terms of the cascading matrix, reduces to the following equalities:

$$\dot{b} = \dot{c}$$

and

$$\det \mathbf{A} = 1$$

Among the two-port junctions shown in Fig. 4.2, the butt joint of two transmission lines is antimetric, and all the other elementary two-ports shown in the same figure are symmetrical.

Example 4.1. A single-step transformer (Fig. 4.3). As follows from the principle of decomposition, the cascading matrix of this trans-

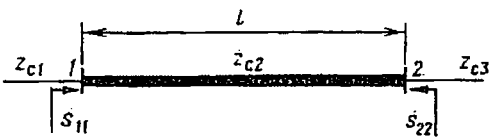


Fig. 4.3. Impedance transformer in a microwave transmission system

former can be obtained on multiplying the A matrix of the left-hand jump in the characteristic impedance by the cascading matrix of the transmission-line section of length l and again by the A matrix of the right-hand jump in the characteristic impedance. This gives

$$\mathbf{A} = \begin{bmatrix} (z_{c3}/z_{c1})^{1/2} \cos \beta l & j\,[z_{c2}/(z_{c1}z_{c3})^{1/2}] \sin \beta l \\ j\,[(z_{c1}z_{c3})^{1/2}/z_{c2}] \sin \beta l & (z_{c1}/z_{c3})^{1/2} \cos \beta l \end{bmatrix} \tag{4.6}$$

Applying the conversion equations to the scattering matrix in (4.5), we obtain the input reflection coefficient of the transformer

$$\dot{s}_{11} = \frac{\left[\left(\frac{z_{c3}}{z_{c1}}\right)^{1/2} - \left(\frac{z_{c1}}{z_{c3}}\right)^{1/2}\right] \cos \beta l + j \left[\frac{z_{c2}}{(z_{c1}z_{c3})^{1/2}} - \frac{(z_{c1}z_{c3})^{1/2}}{z_{c2}}\right] \sin \beta l}{\left[\left(\frac{z_{c3}}{z_{c1}}\right)^{1/2} + \left(\frac{z_{c1}}{z_{c3}}\right)^{1/2}\right] \cos \beta l + j \left[\frac{z_{c2}}{(z_{c1}z_{c3})^{1/2}} + \frac{(z_{c1}z_{c3})^{1/2}}{z_{c2}}\right] \sin \beta l} \tag{4.7}$$

It can readily be verified that the transformer provides a perfect match, $\dot{s}_{11} = 0$, in the following cases:

(1) at $\beta l = \pi/2$, if $z_{c2} = (z_{c1}z_{c3})^{1/2}$. This is an ordinary quarter-wave transformer. Note that it is antimetric, that is, $\dot{b} = \dot{c}$;

(2) at $\beta l = \pi$ and any values of z_{c2} if $z_{c1} = z_{c3}$. This is an ordinary half-wave transformer. It is symmetrical because $\dot{a} = \dot{d}$;

(3) at any βl if $z_{c1} = z_{c2} = z_{c3}$. This is the trivial case of a uniform transmission line.

By introducing a small amount of detuning in terms of wavelength, such that $\Delta\lambda = \lambda - \lambda_0$, and expanding the trigonometric functions in (4.7) into a Taylor series in the vicinity of the nominal wavelength

λ_0, we can deduce simple approximate equations with which to estimate the frequency behaviour of the transformer with the given mismatch:

Quater-wave transformer	Half-wave transformer
$l = \lambda_0/4, \quad z_{c2}^2 = z_{c1} z_{c2}$	$l = \lambda_0/2, \quad z_{c1} = z_{c3}$
$\beta l \approx \frac{2\pi l}{\lambda_0} \frac{\lambda_0}{\lambda} \approx \frac{\pi}{2}(1 - \Delta\lambda/\lambda_0)$	$\beta l \approx \pi (1 - \Delta\lambda/\lambda_0)$
$\sin \beta l \approx 1, \cos \beta l \approx (\pi/2)(\Delta\lambda/\lambda_0)$	$\sin \beta l \approx \pi (\Delta\lambda/\lambda_0), \cos \beta l \approx -1$
$\dot{s}_{11} \approx j \frac{\pi}{4} \left(\sqrt{\frac{z_{c1}}{z_{c3}}} - \sqrt{\frac{z_{c3}}{z_{c1}}} \right) \frac{\Delta\lambda}{\lambda_0}$	$\dot{s}_{11} \approx j \frac{\pi}{2} \left(\frac{z_{c1}}{z_{c2}} - \frac{z_{c2}}{z_{c1}} \right) \frac{\Delta\lambda}{\lambda_0}$

The linearization of trigonometric functions used in deriving the above equations in the case of a small amount off-tune will likewise be convenient in analysis of other devices containing resonant transmission-line sections.

4.3. Decomposition of Symmetrical Four-Ports (the Even and Odd Excitation Method)

With this approach, analysis of four-ports having a plane of symmetry is reduced to that of simpler two-ports which are 'halves' of the original four-port.

Let the plane of symmetry P_1 of a reciprocal four-port pass between ports *1* and *3*, on the one hand, and ports *2* and *4*, on the other

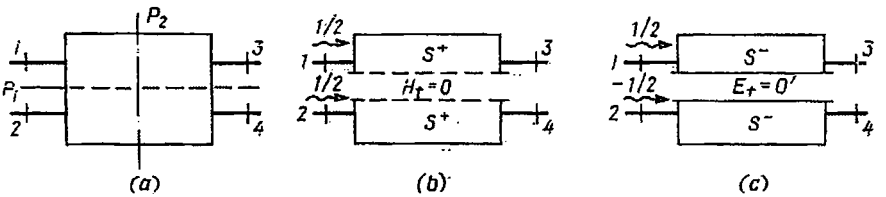


Fig. 4.4. Four-port decomposition: *a*, *b*, the even-excitation case; *c*, the odd-excitation case

(Fig. 4.4*a*). Owing to symmetry and reciprocity, only six elements in the scattering matrix of the four-port are independent: they are $\dot{s}_{11} = \dot{s}_{22}$, $\dot{s}_{21} = \dot{s}_{12}$, $\dot{s}_{31} = \dot{s}_{13} = \dot{s}_{42} = \dot{s}_{24}$, $\dot{s}_{41} = \dot{s}_{14} = \dot{s}_{23} = \dot{s}_{32}$, $\dot{s}_{33} = \dot{s}_{44}$, and $\dot{s}_{43} = \dot{s}_{34}$. Consider the operation of the four-port in two excitation modes: even and odd.

1. With *even excitation* (that is, equal and in phase) applied to ports *1* and *2* in the form of incident waves $\dot{v}_{i1}^{+} = 1/2$ and $\dot{v}_{i2}^{+} = 1/2$, no power will be transferred across the plane of symmetry, and the

plane will contain a loop (an antinode) of a tangential electric field and a node of a tangential magnetic field or, which is the same, an open-circuit condition will result. In a way, the plane of symmetry cuts the four-port into two isolated partial even-excited two-ports, each of which has a scattering matrix, S^+, of its own (Fig. 4.4*b*).

2. With *odd excitation* (that is, equal but in anti-phase) applied to ports *1* and *2* in the form of incident waves $\dot{v}_{i1}^- = 1/2$ and $\dot{v}_{i2}^- = -1/2$, no power will be transferred across the plane of symmetry, and the plane will contain a node of a tangential electric field and a loop of a tangential magnetic field or, which is the same, a short-circuit condition will result. In a way, the plane of symmetry cuts the four-port into two decoupled partial odd-excited two-ports, each having a scattering matrix, S^-, of its own (Fig. 4.4*c*).

The scattering matrices of the even- and odd-excited partial two-ports

$$S^+ = \begin{bmatrix} \dot{\rho}_1^+ & \dot{t}^+ \\ \dot{t}^+ & \dot{\rho}_2^+ \end{bmatrix}, \quad S^- = \begin{bmatrix} \dot{\rho}_1^- & \dot{t}^- \\ \dot{t}^- & \dot{\rho}_2^- \end{bmatrix}$$

may be determined independently of each other, say, in terms of the product of the cascading matrices if the partial two-ports can each be visualized as a cascade of elementary elements.

The superposition of even and odd excitation applied to ports *1* and *2* of a four-port leaves the excitation applied to only one port, *1*:

$$\dot{v}_{i1} = \dot{v}_{i1}^+ + \dot{v}_{i1}^- = 1$$

and

$$\dot{v}_{i2}^+ = \dot{v}_{i2}^+ + \dot{v}_{i2}^- = 0$$

Therefore, on combining the responses of the partial two-ports to even and odd excitation at ports *1* and *2*, it is possible to determine the following four independent scattering coefficients of the four-port:

$$\begin{aligned} \dot{s}_{11} &= \frac{1}{2}(\dot{\rho}_1^+ + \dot{\rho}_1^-) \\ \dot{s}_{21} &= \frac{1}{2}(\dot{\rho}_1^+ - \dot{\rho}_1^-) \\ \dot{s}_{31} &= \frac{1}{2}(\dot{t}^+ + \dot{t}^-) \\ \dot{s}_{41} &= \frac{1}{2}(\dot{t}^+ - \dot{t}^-) \end{aligned} \tag{4.8}$$

In order to find the scattering coefficients $\dot{s}_{33}$ and $\dot{s}_{43}$ of the four-port, it will suffice to apply even and odd excitation to ports *3* and

4. This will lead to the desired relations

$$\dot{s}_{33} = \frac{1}{2}(\dot{\rho}_2^+ + \dot{\rho}_2^-)$$
$$\dot{s}_{43} = \frac{1}{2}(\dot{\rho}_1^+ - \dot{\rho}_2^-) \qquad (4.8a)$$

Naturally, if the four-port has one more plane of symmetry (or all the four ports show a complete rotational symmetry), Eqs. (4.8a) will automatically yield

$$\dot{s}_{33} = \dot{s}_{11}$$

and

$$\dot{s}_{43} = \dot{s}_{21}$$

The even and odd excitation method discussed above may be generalized to symmetric multiports with any even number of ports [2].

Example 4.2. A single-hole sidewall directional coupler. This is a reactive four-port formed by two parallel, rectangular TE_{10}

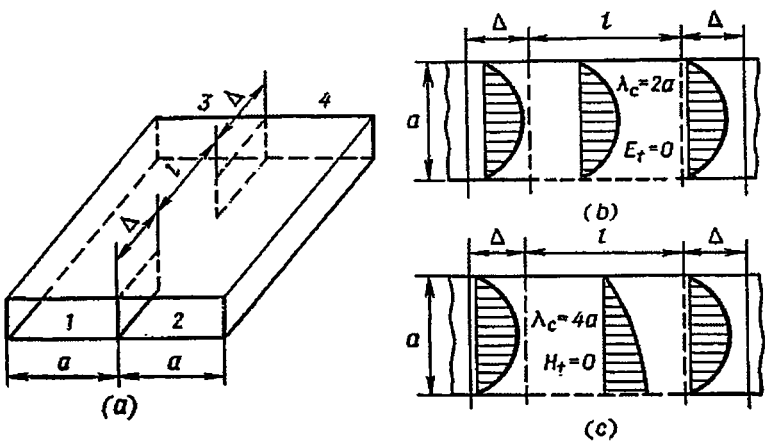


Fig. 4.5. To analysis of (*a*) a single-hole sidewall directional coupler, (*b*) an odd-excited two-port and (*c*) an even-excited two-port

waveguides. The common side-wall (which is a plane of symmetry) contains a coupling hole of length l (Fig. 4.5*a*). Let us locate the terminal planes in all the arms the same distance Δ from the edges of the coupling hole and decompose the junction by applying even and odd excitation to ports *1* and *2*.

The simplest case is that of odd (or antiphase) excitation. It occurs if we mentally stop the coupling hole by an ideal electric wall (Fig. 4.5*b*). The resultant partial odd-excited two-ports are sections of a uniform rectangular waveguide of length $l + 2\Delta$, for

which the scattering parameters are

$$\dot{\rho}_{\bar{1}} = \dot{\rho}_{\bar{2}} = 0$$

and

$$\dot{t}^{-} = \exp(-j\varphi^{-})$$

where

$$\varphi^{-} = (2\pi/\lambda_g^{-})(l + 2\Delta)$$

is the phase delay suffered by the dominant TE_{10} mode in the rectangular waveguide over a distance $l + 2\Delta$, and

$$\lambda_g^{-} = \lambda_0/[1 - (\lambda_0/2a)^2]^{1/2}$$

the wavelength in a guide of width a.

Partial even-excited two-ports result if we mentally stop the coupling hole by an ideal magnetic wall. The middle portion of length l in such a two-port can only sustain the lowest TE mode, with the transverse E and H field components distributed as shown within this portion in Fig. 4.5*c*. Actually, this is a half of the field pattern for the TE_{10} mode in a rectangular metal waveguide of double width $2a$ and with a guide wavelength equal to

$$\lambda_g^{+} = \lambda_0/[1 - \lambda_0/4a)^2]^{1/2}$$

The scattering parameters of an even-excited two-port can be found by solving Maxwell's equations subject to the applicable boundary conditions at the walls. To an approximation well corroborated by a rigorous electrodynamic analysis, it may be deemed that reflections at the joints between an ordinary rectangular waveguide and a hypothetical one having the same cross-section but with a single ideal magnetic sidewall are negligibly small and that $\dot{\rho}_{1,2} \approx 0$ and $\dot{t}^{+} \approx \exp(-j\varphi^{+})$. The phase delay in an even-excited two-port, φ^{+}, differs from that in the case of odd excitation, φ^{-}, by an excess term, $\delta\varphi$, which appears because the guide wavelength is reduced over the distance l due to the boundary condition $H_t = 0$:

$$\varphi^{+} = \varphi^{-} + \delta\varphi$$

where

$$\delta\varphi = 2\pi l\,(1/\lambda_g^{+} - 1/\lambda_g^{-})$$

On substituting the scattering parameters of the partial two-ports thus obtained in Eqs. (4.8), we obtain the scattering matrix of the original four-port

$$\mathbf{S} \approx \left[\begin{array}{c:c} \mathbf{0} & \mathbf{T}_t \\ \hdashline \mathbf{T} & \mathbf{0} \end{array}\right]$$

where

$$\mathbf{T} = \exp\left[-j\left(\varphi^{-} + \delta\varphi/2\right)\right] \begin{bmatrix} \cos(\delta\varphi/2) & -j\sin(\delta\varphi/2) \\ -j\sin(\delta\varphi/2) & \cos(\delta\varphi/2) \end{bmatrix}$$

The fact that the four-port in question has two pairs of match-terminated and decoupled ports is an indication that what we have got is a *directional coupler*. If we choose the length of the coupling hole, l, to be such that the phase difference between the transmission

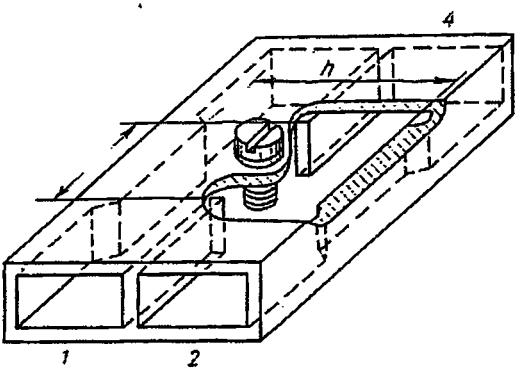


Fig. 4.6. Single-hole sidewall directional coupler

coefficients for the even and odd configurations is $\delta\varphi = \pi/2$, the directional coupler will divide the power entering any of its ports into two equal parts. Thus, this device is a 3-dB aperture-coupled directional coupler. The divided power appears with a *quadrature (90-degree) phase difference* at the output reference planes. This property is characteristic of all directional couplers with two significant planes of symmetry.

In practice, it is usual to insert at the middle of a single-hole sidewall directional coupler a tuning screw adjustable for the depth of immersion, and to reduce the width of the waveguide arms within the coupling length l (Fig. 4.6). This improves the quality of match for the even-excited two-port and enables the power division ratio to be adjusted within narrow limits. With appropriately chosen tuning and other elements, a single-hole sidewall directional coupler will have a bandwidth of 10 to 15% of the centre (operating) frequency.

4.4. Directivity Kinds of Directional Couplers

Directional couplers form a large class of major basic elements used to build branched microwave transmission (feeder) and microwave measurement systems. Most directional couplers have a significant

plane symmetry, so it is advantageous to select the element values and to analyse the associated scattering matrices by the even and odd excitation method, using Eqs. (4.8). Depending on which ports are decoupled, any one of two kinds of directivity may exist.

(1) Directivity of the first kind, when pairs of decoupled ports are *1-2* and *3-4*.

(2) Directivity of the second kind, when pairs of decoupled ports are *1-4* and *2-3*.

In principle, there may be a further kind of directivity in which the decoupled pairs of ports are *1-3* and *2-4*, but this case is, as a rule, equivalent to the directivity of the first kind (with the rotation of the four-port and with its ports re-numbered).

Let us discuss the two kinds of directivity in more detail.

Directivity of the first kind. According to Eqs. (4.8), satisfying simultaneously the requirements for complete match at the ports of a directional coupler, $\dot{s}_{11} = \dot{s}_{22} = 0$, and for decoupling between them, $\dot{s}_{21} = 0$, is equivalent to satisfying the following equalities:

$$\dot{\rho}_1^+ = \dot{\rho}_1^- = \dot{\rho}_2^+ = \dot{\rho}_2^- \equiv 0 \qquad (4.9)$$

$$\dot{t}^{\pm} = \exp\left[-j\left(\theta \pm \delta\varphi/2\right)\right]$$

Or, in words, for directivity of the first kind to take place, the two partial two-ports (one even-excited and the other odd-excited) must be match-terminated and only differ in the phase of the transmission coefficients $\dot{t}^+$ and $\dot{t}^-$. The difference in phase between these coefficients

$$\delta\varphi = \arg \dot{t}^- - \arg \dot{t}^+$$

is called the *differential phase shift* for the modes propagated through the matched partial even- and odd-excited two-ports. Directivity of the first kind is implemented in the single-hole sidewall directional coupler of Example 4.2.

First-kind directional couplers belong to what might be called co-directional couplers because the wave in the secondary transmission line, arms *2-4*, is propagated in the same direction as the wave inducing it and propagated in the primary transmission line, arms *1-3*. Also, first-kind directional couplers are quadrature directional couplers because the phase shift between the output waves of the coupler, with excitation applied to any port, is $\pi/2$, or 90 deg. As a rule, a first-kind directional coupler has a second plane of symmetry (at least, electrical) which passes between the pairs of ports *1-3* and *2-4*.

Equations (4.9) defining directivity of the first kind may be recast in terms of the classical cascading matrices A of the partial two-ports making up the four-port. On invoking the conversion equati-

ons, (4.5), the condition of match for the partial two-ports takes the form

$$(\dot{a}^{\pm} - \dot{d}^{\pm}) + (\dot{b}^{\pm} - \dot{c}^{\pm}) = 0$$

Hence, recalling that in a reactive two-port the coefficients $\dot{a}$ and $\dot{d}$ are real and the coefficients $\dot{b}$ and $\dot{c}$ are imaginaries, it follows that for directivity of the first kind

$$\dot{a}^{\pm} = \dot{d}^{\pm}$$

and

$$\dot{b}^{\pm} = \dot{c}^{\pm}$$

Where necessary, the differential phase shift can also be expressed directly in terms of the elements of the matrices A^{+} and A^{-}.

Directivity of the second kind. In agreement with Eqs. (4.8), the conditions for decoupling, $\dot{s}_{41} = 0$, and for complete match, $\dot{s}_{11} = \dot{s}_{33} = 0$, can be satisfied simultaneously only if $\dot{t}^{+} = \dot{t}^{-}$ and $\dot{\rho}_{1,2}^{+} = -\dot{\rho}_{1,2}^{-}$. In view of the canonical scattering matrix for a reactive two-port, defined by (3.18), this leads to

$$\begin{aligned} \dot{t}^{\pm} &= \sin\tau \exp(j\varphi_2) \\ \dot{\rho}_1^{\pm} &= \pm\cos\tau \exp(j\varphi_1) \\ \dot{\rho}_2^{\pm} &= \mp\cos\tau \exp[j(2\varphi_2 - \varphi_1)] \end{aligned} \tag{4.10}$$

where τ, φ_1 and φ_2 are independent real parameters defining the scattering matrix for a reactive two-port.

If the conditions defined in (4.10) are satisfied, the ideal scattering matrix for a second-kind directional coupler takes the form

$$S^{II} = \left[\begin{array}{cc:cc} 0 & \dot{s}_{21} & \dot{s}_{31} & 0 \\ \dot{s}_{21} & 0 & 0 & \dot{s}_{31} \\ \hdashline \dot{s}_{31} & 0 & 0 & \dot{s}_{43} \\ 0 & \dot{s}_{31} & \dot{s}_{43} & 0 \end{array}\right], \quad \begin{aligned} \dot{s}_{21} &= \dot{\rho}_1^{+} = \cos\tau \exp(j\varphi_1) \\ \dot{s}_{31} &= \dot{t}^{+} = \sin\tau \exp(j\varphi_2) \\ \dot{s}_{43} &= \dot{\rho}_2^{+} = -\cos\tau \exp[j(2\varphi_2 - \varphi_1)] \end{aligned} \tag{4.11}$$

Second-kind directional couplers are *reverse-coupling* directional couplers (or Schwinger couplers) because the wave in the secondary transmission line, arms *2-4*, propagates counter to the inducing wave propagating in the primary transmission line, arms *1-3*. If a second-kind directional coupler has a second plane of symmetry passing between the two pairs of ports *1-2* and *3-4*, then $\dot{\rho}_1^{\pm} = \dot{\rho}_2^{\pm}$, which

implies that $\dot{s}_{43} = \dot{s}_{21}$ and $\varphi_2 = \varphi_1 \pm \pi/2$. Thus, if it has two planes of symmetry, a second-kind directional coupler introduces a *quadrature* phase shift. If, on the other hand, a second plane of symmetry is not there, but the parameters of the partial two-ports can be adjusted so that $\varphi_1 = \varphi_2$, then a second-kind directional coupler reduces to an even-odd-excited device for which $\dot{s}_{43} = -\dot{s}_{21}$.

Let us express the conditions for directivity of the second kind in terms of the elements of the A matrices for the component two-ports. Expanding (4.10) with the aid of the equations (4.5) for conversion from the elements of the A matrix to those of the S matrix, we obtain the following equalities:

$$(\dot{a}^+ + \dot{d}^+) + (\dot{b}^+ + \dot{c}^+) = (\dot{a}^- + \dot{d}^-) + (\dot{b}^- + \dot{c}^-)$$

$$(\dot{a}^+ - \dot{d}^+) + (\dot{b}^+ - \dot{c}^+) = -(\dot{a}^- - \dot{d}^-) - (\dot{b}^- - \dot{c}^-)$$

Since $\dot{a}$ and $\dot{d}$ are real while $\dot{b}$ and $\dot{c}$ are imaginaries, the condition for directivity of the second kind takes a very simple form

$$\dot{a}^\pm = \dot{d}^\mp \text{ and } \dot{b}^\pm = \dot{c}^\mp \tag{4.12}$$

Using the above conditions for directional couplers expressed in terms of the parameters of the partial two-ports, it is possible not only to get insight into the operation of a coupler, but also to derive the required design relations.

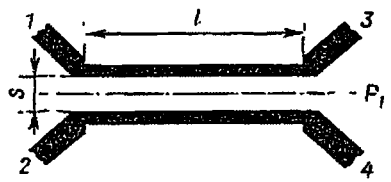


Fig. 4.7. Conductor topology of a directional coupler

Example 4.3. A directional coupler constructed in coupled TEM-mode transmission lines (Fig. 4.7). This is the most compact and broadband type of directional coupler. One of the planes of symmetry passes between strip-line conductors. Relative to this plane of symmetry the partial even- and odd-excited two-ports take the form of TEM-mode uniform line sections whose length, l, is the same as that of the connecting section, and their characteristic impedances $z_c^\pm$ depend on which boundary condition ($H_t = 0$ or $E_t = 0$) exists in the plane of symmetry.

The cross-sections and the transverse electric field patterns of the partial even- and odd-excited two-ports for three forms of coupled transmission lines are shown in Fig. 4.8. As follows from comparison of the electric field patterns, the capacitance per unit of line length in the even-excited case ($H_t = 0$) must be smaller than that in the odd-excited case ($E_t = 0$). Hence, the characteristic impedances of the partial two-ports satisfy the following inequality

$$z_c^+ > z_c^-$$

Assuming that the leads of the directional coupler have a normalized characteristic impedance of unity, we may, in agreement with Eq. (4.6) write the classical cascading matrices for the individual two-ports as follows:

$$\mathbf{A}^{\pm} = \begin{bmatrix} \cos\beta^{\pm} l & (jz_c^{\pm}) \sin\beta^{\pm} l \\ (j/z_c^{\pm}) \sin\beta^{\pm} l & \cos\beta^{\pm} l \end{bmatrix}$$

With the above form of individual two-ports we can only implement directivity of the second kind, provided that the elements of the matrices $\mathbf{A}^{\pm}$ satisfy the conditions defined in (4.12). Under these conditions, the equality $\dot{a}^{\pm} = \dot{d}^{\mp}$ will be satisfied only if the two

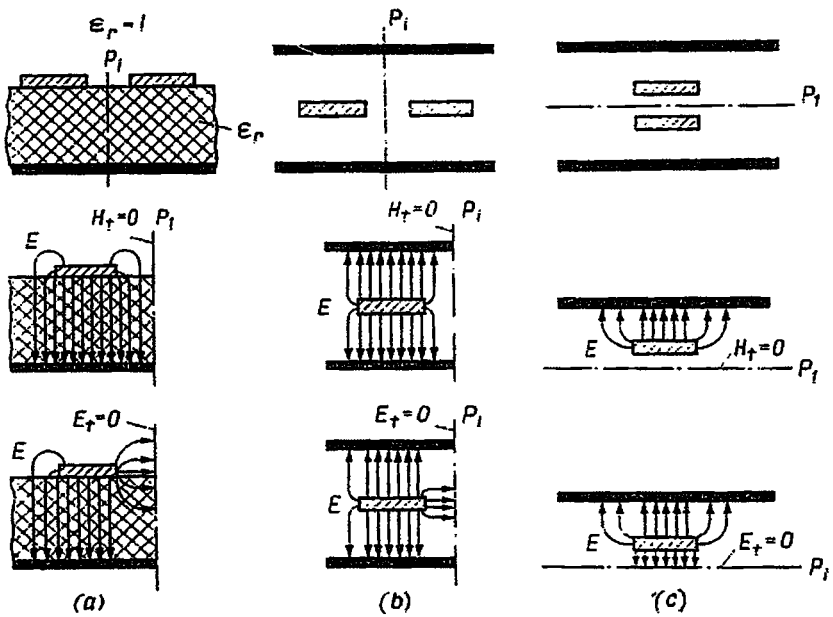


Fig. 4.8. Cross-sections of coupled microstrip and strip lines and the respective cross-sections of constituent even-excited and odd-excited two-ports: *a*, side-coupled microstrip lines; *b*, side-coupled strip lines; *c*, face-coupled strip lines

individual two-ports have the same phase constants, $\beta^{+} = \beta^{-} = \beta$, and the equality $\dot{b}^{\pm} = \dot{c}^{\mp}$ will be satisfied if the characteristic impedances are chosen such that $z_c^{+} z_c^{-} = 1$.

Importantly, the directivity conditions thus derived are independent of the electrical length βl of the connecting line, and this impli-

es that *ideal decoupling between and a complete match at the ports will be obtained at any frequency*. The only frequency-dependent parameter is the power distribution at the coupler ports. Using Eqs. (4.11) and the rules for conversion from the cascading matrix to the scattering matrix, and carrying out simple manipulations subject to the condition that $z_c^+ z_c^- = 1$, we get

$$\dot{s}_{21} = \frac{jM \sin \beta l}{(1-M^2)^{1/2} \cos \beta l + j \sin \beta l}$$
$$\dot{s}_{31} = \frac{(1-M^2)^{1/2}}{(1-M^2)^{1/2} \cos \beta l + j \sin \beta l} \qquad (4.13)$$

where

$$M = (z_c^+ - z_c^-)/(z_c^+ + z_c^-)$$

is the coupling factor.

If the length of the coupling section is an odd multiple of a quarter-wavelength in the line, the transmission coefficient $\dot{s}_{21}$ is a maximum and equal to the parameter M. If, on the other hand, the length of the coupling section is a multiple of a half-wavelength, then $\dot{s}_{21} = 0$, and all of the power entering port *1* will be transferred to port *3*. The manner in which the magnitudes of s_{21} and s_{31} vary with the electrical length βl of a coupled-line directional coupler for $M = 1/\sqrt{2}$ is shown in the plot of Fig. 4.9.

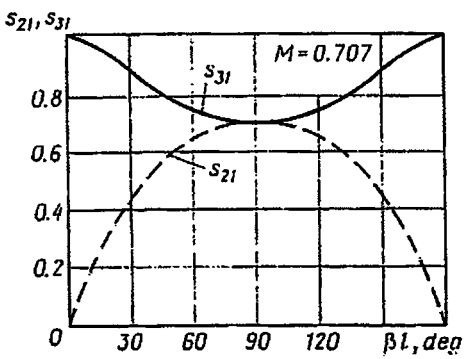


Fig. 4.9. Frequency response of a 3-dB coupled-line directional coupler

Coupled-line directional couplers have two significant planes of symmetry, so they belong to *quadrature* directional couplers.

When designing a directional coupler for which the transmission coefficients have been specified in advance such that

$$|\dot{s}_{21}|^2 + |\dot{s}_{31}|^2 = 1$$

the length of the coupling section may be chosen at will (most often, equal to a quarter-wavelength in the guide). Then, using Eqs. (4.13), one can find the parameter M and determine the characteristic impedances z_c^+ and z_c^-. Using these impedances as the point of departure, one can then calculate the required dimensions of strip conductors in the cross-section of the directional coupler and the spacing between them. At this stage, resort has to be made either to reference sources on the characteristic impedances of coupled transmissi-

on lines, or to computer programs compiled with special reference to the chosen cross-sectional geometry of the coupled lines. It is most difficult to build directional couplers with a large amount of coupling ($M > 0.3$) because the spacing s between the conductors (see Fig. 4.7) must be extremely small, thus posing manufacturing problems. Several techniques have been developed which offer an opportunity to maintain a large amount of coupling with the specified

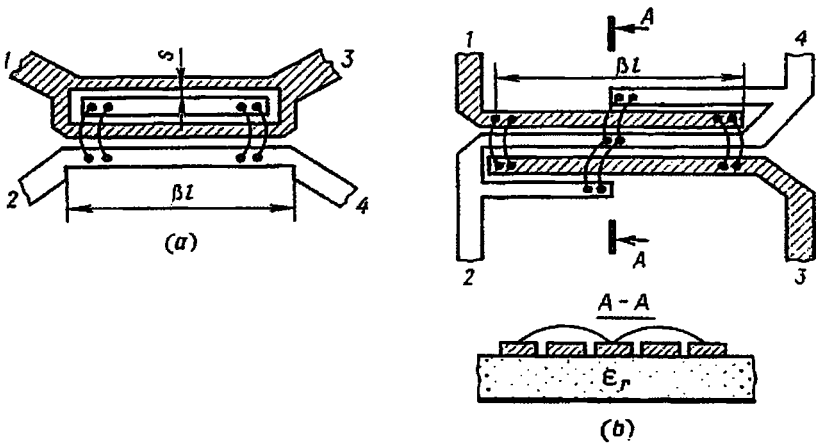


Fig. 4.10. Coupled-line directional couplers due to Lange, USA: *a*, with the input lines not crossed; *b*, with the input lines crossed

minimal spacing. One is to build the coupling section from transposed leads as shown in Fig. 4.10. In this type of coupler proposed by Lange of the United States, each coupled line consists of a number of parallel strip lines joined by jumpers. An advantage of the Lange coupler is that it is possible, wherever necessary, to cross the coupled lines. This need may sometimes arise when it is desired to arrange the outputs of a coupler more conveniently for access.

Microstrip directional couplers pose one more problem in their design because the even- and odd-excited waves have different phase constants, β^+ and β^-. The phase velocities of these waves can be equalized by applying an additional dielectric layer to the coupled leads or by inserting reactive compensators at the ends of the coupling section.

4.5. Ring-Type Directional Couplers

Strip, coax and waveguide microwave transmission systems are often built to include ring-type directional couplers which possess both kinds of directivity.

A *square coupler* (Fig. 4.11*a*) consists of four quarter-wave transmission-line sections connected in a square. It implements directivity of the first kind. The input leads of unity characteristic impedance are connected in parallel at the corners of the square.

A *hybrid ring* (*rat-race*), as seen from Fig. 4.11*b*, differs from a square coupler in that the length of the transmission-line section

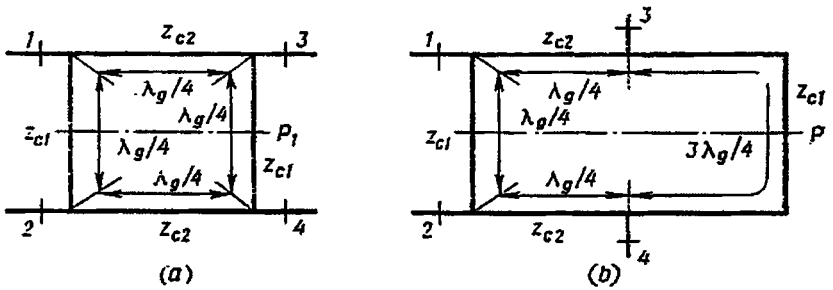


Fig. 4.11. Ring-type directional couplers: *a*, square type; *b*, hybrid ring (rat-race)

between ports *3* and *4* is increased to three-quarters of the guide wavelength, which is equivalent to the addition of a half-wave transformer. The hybrid ring implements directivity of the second kind and is an even-odd-excited coupler.

The variables that permit adjustment of power division at the outputs of ring couplers are the characteristic impedances of the ring sections, z_{c1} and z_{c2}. In the case of a symmetric design, both the square coupler and the hybrid ring may be analysed, using the even-odd excitation method with reference to a horizontal plane. This will yield the following design relations for the elements of the ideal scattering matrices at the centre frequency:

$$1/z_{c2}^2 - 1/z_{c1}^2 = 1$$
$$s_{11} = 0$$
$$s_{21} = s_{34} = 0$$
$$s_{31} = -jz_{c2}$$
$$s_{41} = -z_{c2}/z_{c1}$$

for the square coupler (directivity of the first kind), and

$$1/z_{c2}^2 + 1/z_{c1}^2 = 1$$
$$s_{11} = 0$$
$$s_{14} = s_{23} = 0$$
$$s_{21} = -j/z_{c1}$$

$$\dot{s}_{31} = -j/z_{c2}$$

$$\dot{s}_{43} = -\dot{s}_{21}$$

$$\dot{s}_{42} = \dot{s}_{31}$$

for the hybrid ring (directivity of the second kind).

There is, however, another method which does not take into account the property of symmetry; this is the decomposition of ring

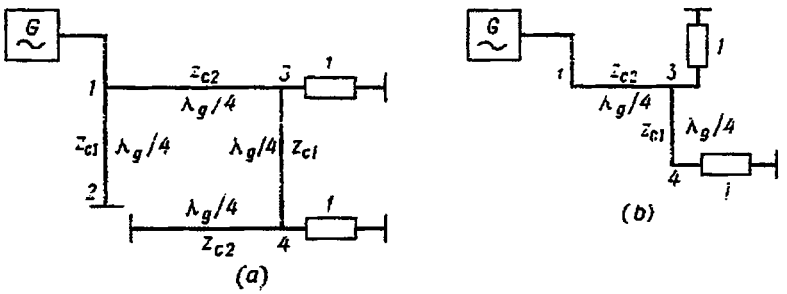


Fig. 4.12. Decomposition of a square directional coupler

couplers. With this approach, the ring in question is mentally broken at a suitable place, and the circuit is reduced to a cascaded connection of several elementary two-ports.

Example 4.4. Decomposition of a square coupler (see Fig. 4.11*a*). With the generator connected to port *1*, it is possible, on the basis of the assumed condition of decoupling ($\dot{s}_{21} = 0$), to obtain a short-circuit at port *2*. As a result, the circuit reduces to the form shown in Fig. 4.12*a*. The shorted quarter-wave stubs thus produced have no shunting effect at ports *1* and *4* and may be omitted. In the simplified circuit (Fig. 4.12*b*), the quarter-wave transformer of characteristic impedance z_{c2} at port *3* will then be loaded into a parallel connection of two conductances, namely $g_3 = 1$ (the outgoing lead) and $g_4 = 1/z_{01}^2$ which is the conductance of outgoing arm *4*, transferred via the quarter-wave transformer z_{c1}. By varying the value of z_{c1}, it is possible to adjust power division at ports *3* and *4* of the coupler because

$$|\dot{s}_{31}|^2/|\dot{s}_{41}|^2 = g_3/g_4 = z_{c1}^2$$

The characteristic impedance of the quarter-wave transformer z_{c2} should be adjusted so as to satisfy the condition for a match at port *1*:

$$1/z_{c2}^2 = 1 + 1/z_{c1}^2$$

The simplest way to determine the scattering coefficient $\dot{s}_{31}$ is from the elements of the A matrix of the quarter-wave transfor-

mer z_{c2}:

$$\dot{c} = j/z_{c2}$$

and

$$\dot{d} = 0$$

(see Fig. 4.2*a*). Assuming that the normalized current at the match-terminated port *1* is

$$\dot{i}_1 = \dot{v}_1 = 1$$

we obtain at once

$$\dot{s}_{31} = \dot{v}_3 = \dot{i}_1/\dot{c} = -jz_{c2}$$

The other non-zero element in the first column of the scattering matrix is

$$\dot{s}_{41} = -z_{c2}/z_{c1}$$

which stems from the specified power division ratio and the unitary property of the scattering matrix.

The assumption that $\dot{s}_{21} = 0$ need not be verified because ports *1* and *2* are decoupled in view of the match already secured at the

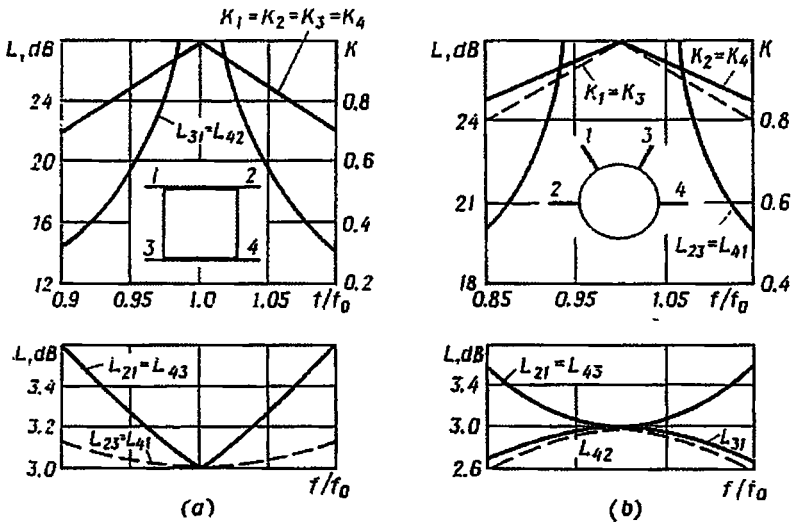


Fig. 4.13. Frequency response of (*a*) a square directional coupler and (*b*) a hybrid ring

ports of the coupler (see the theorem in Example 3.4). Note that the hybrid ring and many other types of ring couplers can be decomposed in a similar way.

The condition of decoupling and match obtained by adjustment of characteristic impedances can be secured for ring couplers only at the design frequency; they are violated at any other frequency. It is customary to state the performance of directional couplers over the bandwidth in terms of the following parameters: the input TWR

$$K_i = (1 - |\dot{s}_{ii}|)/(1 + |\dot{s}_{ii}|)$$

and the coupling factor (in dB)

$$L_{ij} = -20 \log |\dot{s}_{ij}|$$

where i and j mark the ith and jth ports, respectively. The design frequency responses of a square coupler and a hybrid ring are shown in Fig. 4.13 which also illustrates the likely topology of printed-circuit boards.

The results have been obtained, using a universal interactive program, MAKET, compiled by B.A. Mishustin for determining the scattering matrices of complex electronic circuits. The square coupler has a rather limited bandwidth which is not over $\pm 5\%$ as compared with $\pm 15\%$ for the hybrid ring.

4.6. Matched Three-Port Power Dividers

Power dividers are major basic elements used to split power among the arms of a branched microwave transmission system. A three-port power divider has a main (input) arm or port, *1*, and two output arms (or ports), *2* and *3*. As a rule, a power divider is required to satisfy the match condition at the main port, $\dot{s}_{11} = 0$, and the condition for power transfer from the input port to the output port in such a way that the transmission coefficients $\dot{s}_{21}$ and $\dot{s}_{31}$ have specified magnitudes. Power dividers can also serve to combine in arm *1* the power flows from two coherent sources connected to ports *2* and *3*. In this and some other applications, a power divider is to satisfy the additional conditions for match and decoupling between ports *2* and *3*.

We have already discussed a simple T-junction (see Fig. 3.2). It belongs to the class of reactive three-port junctions, so, according to the theorem in Example 3.4, it cannot be matched at all the three ports at the same time. In order to secure a match and decoupling between ports *2* and *3* of a three-port power divider, it is necessary to insert absorbing elements in its circuit. The most common design of what is known as an equal power divider is shown in Fig. 4.14. The divider uses a parallel junction of transmission li-

nes at port *1*, two quarter-wave transformers with a characteristic impedance $z_c = \sqrt{2}$, and an absorber in the form of a lumped re-

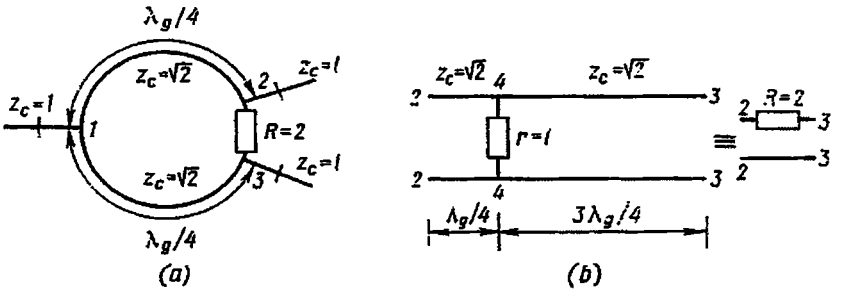


Fig. 4.14. Matched ring-type equal-power divider: *a*, general arrangement; *b*, equivalent transformation of part of the hybrid ring between ports *2-4-3*

sistor with a normalized resistance equal to $R = 2$. This design can be derived from a hybrid ring (see Fig. 4.11*b*). The hybrid-ring segments between ports *2* and *3* are shown separately in Fig. 4.14*b*. The A matrix of this section has the form

$$\mathbf{A} = \begin{bmatrix} 0 & j\sqrt{2} \\ j/\sqrt{2} & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \times \begin{bmatrix} 0 & -j\sqrt{2} \\ -j/\sqrt{2} & 0 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$$

That is, it is identical to the matrix of the series lumped resistor ($R = 2$) placed directly between ports *2* and *3* (see Fig. 4.2).

Thus, the *matched power divider* shown in Fig. 4.14*a* is in effect a *directional coupler* in which the role of port *4* is played by a resistor ($R=2$). Note that in terms of symmetry this power divider is analogous to the magic-tee shown in Fig. 3.8.

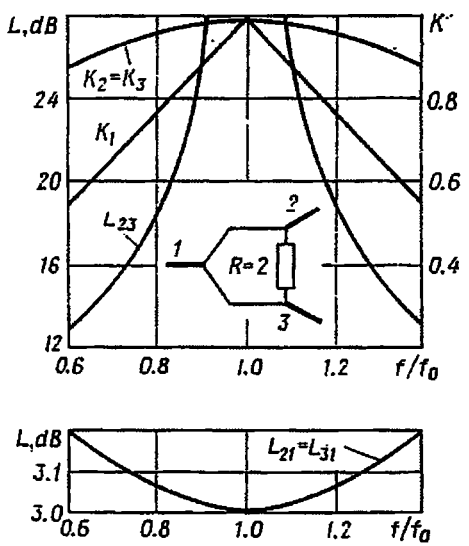


Fig. 4.15. Frequency response of a matched power divider

The performance of a power divider over the bandwidth is stated in terms of the TWR of each port, K_t, and the coupling factor (in dB), $L_{ij} = -20 \log |\dot{s}_{ij}|$. The design frequency dependences of the parameters are shown in Fig. 4.15. The operating bandwidth is

about ±20%, which is somewhat greater than it is in the case of a hybrid ring because the frequency-dependent ring segment between ports *2* and *3* is replaced by a frequency-independent resistor.

The likely designs of unequal power dividers are shown in Fig. 4.16. The design relations for the elements of the ideal scatte-

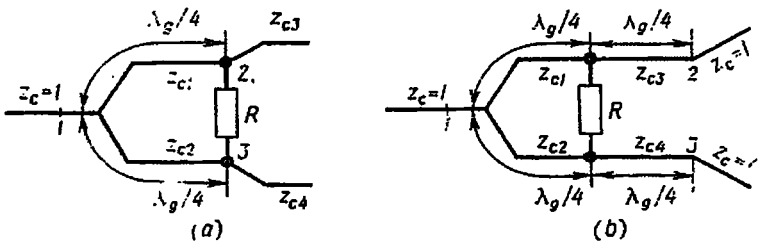


Fig. 4.16. Matched three-port power dividers: *a*, with output lines differing in characteristic impedance; *b*, with matching transformers in output lines

ring matrices of power dividers at the centre frequency have the form

$$\begin{aligned} s_{11} = s_{22} = s_{33} &= 0 \\ \dot{s}_{21} &= -jt \\ \dot{s}_{31} &= -j\,(1 - t^2)^{1/2} \\ t &= (z_{c3}/z_{c1})^{1/2} \\ (1 - t^2)^{1/2} &= (z_{c4}/z_{c2})^{1/2} \\ R &= z_{c3} + z_{c4} = z_{c1}z_{c2} \end{aligned}$$

for a divider with modified characteristic impedances of the outgoing leads (Fig. 4.16*a*), and

$$\begin{aligned} s_{11} = s_{22} = s_{33} &= 0 \\ \dot{s}_{21} &= -t \\ \dot{s}_{31} &= -(1 - t^2)^{1/2} \\ t &= z_{c3}/z_{c1} \\ (1 - t^2)^{1/2} &= z_{c4}/z_{c2} \\ R &= z_{c3}^2 + z_{c4}^2 = z_{c1}z_{c2} \end{aligned}$$

for a power divider with matching transformers in the outgoing leads (Fig. 4.16*b*).

One of the techniques to deduce the above relations on the basis of the even-odd excitation method is set forth in [2].

4.7. The Scattering Matrix of Cascaded Multiport Junctions

In this section, we shall develop equations leading to the scattering matrix of a cascaded connection of two multiports each having a scattering matrix known in advance. These equations are crucial in the theory of microwave circuits and offer an opportunity to compile computer programs to find scattering matrices for any linear network by consecutively stringing the individual basic elements. Using these equations, one can also present the specified ideal scattering matrix of a multiport junction as a result of com-

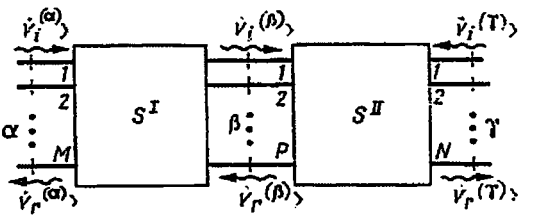


Fig. 4.17. Cascade connection of two multiports

bining the scattering matrices of several simpler devices. In this way, one can readily synthesize a multiport junction from its specified scattering matrix.

The most common form of the cascaded connection of two multiports, applicable to a multiplicity of practical designs, is shown in Fig. 4.17. Each multiport junction has a group of ports not involved in the connection (the first multiport has a number M of such ports, and the second, N ports). If the second multiport has no 'free' ports ($N = 0$), it will act simply as a multi-input load for the first multiport.

To obtain the resultant scattering matrix S^{Σ} for a cascaded connection of multiports, it is advisable to number the ports in a certain definite manner. To begin with, let us number all the M free ports of multiport I (the group of ports labelled α) and the P ports engaged in the connection (the group of ports labelled β). In multiport II, we shall first number the P ports engaged in cascading (likewise labelled β), then the N free ports (labelled γ). With this numbering scheme, we can split each of the matrices S^I and S^{II} into four blocks in terms of which the results obtained by combi-

ning the multiports will be expressed. So, let us write the scattering matrices of the original multiports as follows:

$$S^{I} = \left[\begin{array}{c|c} S_{\alpha\alpha} & S_{\alpha\beta} \\ \hline S_{\beta\alpha} & S^{I}_{\beta\beta} \end{array}\right]$$

$$S^{II} = \left[\begin{array}{c|c} S^{II}_{\beta\beta} & S_{\beta\gamma} \\ \hline S_{\gamma\beta} & S_{\gamma\gamma} \end{array}\right]$$

The S^{I} matrix is a square one of order $M+P$. The blocks $S_{\alpha\alpha}$ and $S^{I}_{\beta\beta}$ are likewise square matrices of order M and P, respectively. In the general case, the off-diagonal blocks are rectangular: the block $S_{\alpha\beta}$ is an $M \times P$ matrix (it has M rows and P columns), and the block $S_{\beta\alpha}$ is a $P \times M$ matrix. S^{II} is a $P+N$ matrix, and its diagonal blocks $S^{II}_{\beta\beta}$ and $S_{\gamma\gamma}$ have the dimension P and N, respectively. The block $S_{\beta\gamma}$ is a $P \times N$ matrix, and the block $S_{\gamma\beta}$ is an $N \times P$ matrix.

As is usual, we will, in port groups α and γ, distinguish incident waves travelling towards the junction and reflected waves receding from it. In port group β, the waves in the transmission lines connecting the multiports cannot be identified as incident or reflected. Therefore, we shall use the symbol $v^{\beta}_{12}\rangle$ for the column matrix of the voltage waves travelling towards the first multiport, and the symbol $v^{\beta}_{21}\rangle$ for the voltage waves travelling towards the second multiport. Using the blocks of the S^{I} and S^{II} scattering matrices, the incident and reflected voltage waves can be connected by relations of the form

$$\begin{aligned}
v^{\alpha}_{r}\rangle &= S_{\alpha\alpha}v^{\alpha}_{i}\rangle + S_{\alpha\beta}v^{\beta}_{12}\rangle \\
v^{\beta}_{21}\rangle &= S_{\beta\alpha}v^{\alpha}_{i}\rangle + S^{I}_{\beta\beta}v^{\beta}_{12}\rangle \\
v^{\beta}_{12}\rangle &= S^{II}_{\beta\beta}v^{\beta}_{21}\rangle + S_{\beta\gamma}v^{\gamma}_{i}\rangle \\
v^{\gamma}_{r}\rangle &= S_{\gamma\beta}v^{\beta}_{21}\rangle + S_{\gamma\gamma}v^{\gamma}_{i}\rangle
\end{aligned}$$

If we delete the column matrices of the travelling voltage waves, $v^{\beta}_{21}\rangle$ and $v^{\beta}_{12}\rangle$ from the above matrix equations, the remaining two equations will define the scattering matrix of the cascaded multiport junctions, S^{Σ}:

$$\begin{aligned}
v^{\alpha}_{r}\rangle &= S^{\Sigma}_{\alpha\alpha}v^{\alpha}_{i}\rangle + S_{\alpha\gamma}v^{\gamma}_{i}\rangle \\
v^{\gamma}_{r}\rangle &= S_{\gamma\alpha}v^{\alpha}_{i}\rangle + S^{\Sigma}_{\gamma\gamma}v^{\gamma}_{i}\rangle
\end{aligned} \quad \rightarrow S^{\Sigma} = \left[\begin{array}{c|c} S^{\Sigma}_{\alpha\alpha} & S_{\alpha\gamma} \\ \hline S_{\gamma\alpha} & S^{\Sigma}_{\gamma\gamma} \end{array}\right]$$

in which the blocks of the S^{Σ} scattering matrix are defined as follows:

$$\left.\begin{aligned} S^{\Sigma}_{\alpha\alpha} &= S_{\alpha\alpha} + S_{\alpha\beta}\,(E - S^{II}_{\beta\beta}S^{I}_{\beta\beta})^{-1}S^{II}_{\beta\beta}S_{\beta\alpha} \\ S_{\alpha\gamma} &= S_{\alpha\beta}\,(E - S^{II}_{\beta\beta}S^{I}_{\beta\beta})^{-1}S_{\beta\gamma} \\ S_{\gamma\alpha} &= S_{\gamma\beta}\,(E - S^{I}_{\beta\beta}S^{II}_{\beta\beta})^{-1}S_{\beta\alpha} \\ S^{\Sigma}_{\gamma\gamma} &= S_{\gamma\gamma} + S_{\gamma\beta}\,(E - S^{I}_{\beta\beta}S^{II}_{\beta\beta})^{-1}S^{I}_{\beta\beta}S_{\beta\gamma} \end{aligned}\right\} \qquad (4.14)$$

When using Eqs. (4.14), it is important not to overlook the fact that the *pairwise connected ports* are the *terminal planes belonging to the same transmission line.* This implies that different transmission lines cannot be interconnected directly. A junction of different transmission lines is a separate entity for which a separate parameter matrix must be written and taken into consideration.

On a computer, Eqs. (4.14) can be used to calculate the scattering matrix for any cascade connection of multiports over the bandwidth. The usual procedure is to program the cascading of the first two multiports and, after the calculations have been done, to write the result in the computer memory instead of the scattering matrix for the first multiport. This sequence can be extended to include a third, a fourth multiport, and so on.

For some cascaded connections of multiports Eqs. (4.14) can markedly be simplified and used to derive analytical expressions for the scattering coefficients of major microwave circuits. Consider some typical cases.

1. One of the individual multiports in the cascaded connection is matched and decoupled at all of the connected ports. In this case, the equations are especially simple because there is no need to take inverses of the matrices in Eqs. (4.14). If a complete match and decoupling exist for port group β in either the first or the second multiport, Eqs. (4.14) take the following form:

— for $S^{I}_{\beta\beta} = 0$,

$$\left.\begin{aligned} S^{\Sigma}_{\alpha\alpha} &= S_{\alpha\alpha} + S_{\alpha\beta}S^{II}_{\beta\beta}S_{\beta\alpha} \\ S_{\alpha\gamma} &= S_{\alpha\beta}S_{\beta\gamma} \\ S_{\gamma\alpha} &= S_{\gamma\beta}S_{\beta\alpha} \\ S^{\Sigma}_{\gamma\gamma} &= S_{\gamma\gamma} \end{aligned}\right\} \qquad (4.15a)$$

— for $S^{II}_{\beta\beta} = 0$,

$$\left.\begin{aligned} S^{\Sigma}_{\alpha\alpha} &= S_{\alpha\alpha} \\ S_{\alpha\gamma} &= S_{\alpha\beta}S_{\beta\gamma} \\ S_{\gamma\alpha} &= S_{\gamma\beta}S_{\beta\alpha} \\ S^{\Sigma}_{\gamma\gamma} &= S_{\gamma\gamma} + S_{\gamma\beta}S^{I}_{\beta\beta}S_{\beta\gamma} \end{aligned}\right\} \qquad (4.15b)$$

A distinction of the above equations is that the off-diagonal blocks of the scattering matrix of the cascaded multiport junction, called the transfer (or transmission) blocks, may be calculated as the products of the respective blocks of the scattering matrices of the individual multiport junctions being cascaded.

2. The scattering matrix of a multiport junction with one of its ports terminated in a specified load. In this case, it is convenient to use the general equations (4.14) for the cascaded connection of multiports in which the scattering matrix S^{II} consists of only one block of order one, which is simply the load reflection coefficient $\dot{\rho}_L$. The block

$$S^{I}_{\beta\beta} = \dot{s}_{pp}$$

of the scattering matrix is likewise of order one (here p is the index of the load-terminated port). The blocks $S_{\alpha\beta}$ and $S_{\beta\alpha}$ are respectively the pth column and the pth row scattering matrix for the terminated multiport (of course, with the diagonal element $\dot{s}_{pp}$ dropped). The block $S_{\alpha\alpha}$ is the scattering matrix part that remains upon removal of the elements $\dot{s}_{pp}$, the column $s_{\alpha\beta}\rangle$, and the row $\langle s_{\beta\alpha}$.

Referring to the first line in (4.14) and substituting in it the respective blocks from the scattering matrices, we get

$$S^{\Sigma}_{\alpha\alpha} = S_{\alpha\alpha} + \frac{\dot{\rho}_L}{1 - \dot{s}_{pp}\dot{\rho}_L} s_{\alpha\beta} \rangle \langle s_{\beta\alpha} \qquad (4.16)$$

Thus, when one of the ports of a multiport is terminated in a specified load, the order of the matrix goes down by one, and the scattering matrix itself can be evaluated as the sum of two terms. The first term is the part of the scattering matrix remaining upon removal of the pth column and the pth row, and the second term is a square matrix of the same order as the block $S_{\alpha\alpha}$ and is formed by taking the product of the removed column and row (minus the diagonal element) and multiplying it by the term $\dot{\rho}_L/(1 - \dot{s}_{pp}\dot{\rho}_L)$.

If the multiport in question is a two-port junction, Eq. (4.16) reduces to the ordinary equation for the input reflection coefficient of a two-port terminated in a specified load

$$\dot{\rho}_{in} = \dot{s}_{11} + \dot{s}_{12}\dot{s}_{21}\dot{\rho}_L/(1 - \dot{s}_{22}\dot{\rho}_L) \qquad (4.17)$$

When several ports of a multiport are terminated in several loads, its scattering matrix can be transformed consecutively, using Eq. (4.16) and reducing the order of the scattering matrix by one each time. Note that the resultant scattering matrix will lose its unitary property if absorbing loads are used as terminations for a lossless multiport.

4.8. Examples of Cascaded Four- and Two-Ports

Example 4.5. A cascaded connection of two directional couplers (Fig. 4.18). The directional couplers are constructed in TEM-mode coupled transmission line (see Sec. 4.4) and their scattering matrices are

$$\mathbf{S}^{\mathrm{I,\,II}} = \left[\begin{array}{c|c} \mathbf{0} & \mathbf{T}_t^{\mathrm{I,\,II}} \\ \hline \mathbf{T}^{\mathrm{I,\,II}} & \mathbf{0} \end{array}\right]$$

$$\mathbf{T}^{\mathrm{I,\,II}} = \begin{bmatrix} \sin\tau_{1,2} & -j\cos\tau_{1,2} \\ -j\cos\tau_{1,2} & \sin\tau_{1,2} \end{bmatrix}$$

If the coupled lines are quarter-wave long, the parameter M of each directional coupler will, in agreement with Eqs. (4.13), be

$$M_i = (z_{ci}^{+} - z_{ci}^{-})/(z_{ci}^{+} + z_{ci}^{-}) = \sin\tau_i$$

Cascading does not upset matching and decoupling and produces a *composite directional coupler*. According to Eqs. (4.15), its scattering matrix has the form

$$\mathbf{S}^{\Sigma} = \left[\begin{array}{c|c} \mathbf{0} & \mathbf{T}_t^{\Sigma} \\ \hline \mathbf{T}^{\Sigma} & \mathbf{0} \end{array}\right]$$

$$\mathbf{T}^{\Sigma} = \mathbf{T}^{\mathrm{I}}\mathbf{T}^{\mathrm{II}} = \begin{bmatrix} -\cos(\tau_1+\tau_2) & -j\sin(\tau_1+\tau_2) \\ -j\sin(\tau_1+\tau_2) & -\cos(\tau_1+\tau_2) \end{bmatrix}$$

If each directional coupler in the circuit of Fig. 4.18 is such that $\tau_1 = \tau_2 = \pi/8$ (the attenuation between the primary and secondary lines of the coupler is $L = -20\log|\sin\tau_i| = 8.34$ dB), then the cascading produces an equal power divider, or a 3-dB directional coupler:

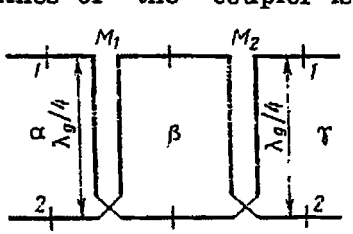


Fig. 4.18. Composite directional coupler

$$\sin(\tau_1+\tau_2) = \cos(\tau_1+\tau_2) = 1/\sqrt{2}$$

Thus, a directional coupler with a large amount of coupling can be obtained by cascading two directional couplers with a small amount of coupling. With loose coupling, there is no need for a prohibitively narrow spacing between the coupled leads, and this substantially simplifies their manufacture.

A composite directional coupler may also consist of square and directional couplers implementing directivity of the first kind. In this case, an added advantage is the *expansion of the operating bandwidth* owing to the cancellation of reflections in the cascaded even- and odd-excited individual two-ports.

Example 4.6. A variable-coupling directional coupler. This is a cascaded connection of three simpler networks (Fig. 4.19). Let the first and third networks be each a 3-dB quadrature directional coupler for which the scattering matrices are as follows:

$$\mathbf{S}^{\mathrm{I,\,II}} = \left[\begin{array}{c|c} \mathbf{0} & \mathbf{T}_t \\ \hline \mathbf{T} & \mathbf{0} \end{array}\right]$$

$$\mathbf{T} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -j \\ -j & 1 \end{bmatrix}$$

The middle network is a combination of two uncoupled two-ports and has a scattering matrix of the form

$$\mathbf{S}^{\mathrm{II}} = \left[\begin{array}{c|c} \mathbf{0} & \mathbf{T}_t^{\mathrm{II}} \\ \hline \mathbf{T}^{\mathrm{II}} & \mathbf{0} \end{array}\right]$$

$$\mathbf{T}^{\mathrm{II}} = \begin{bmatrix} \exp(-j\varphi) & 0 \\ 0 & 1 \end{bmatrix}$$

where the parameter φ is the excess phase delay in the upper channel due to an adjustable phase-shifter matched at both

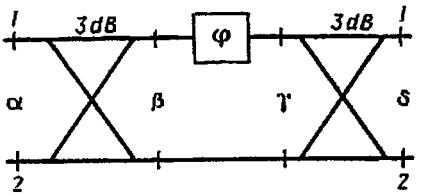


Fig. 4.19. Variable-coupling directional coupler

ends. Applying twice Eqs. (4.15) for the cascading of matched and decoupled multiports, we obtain the following scattering matrix for the cascade connection of the three individual networks:

$$\mathbf{S}^{\Sigma} = \left[\begin{array}{c|c} \mathbf{0} & \mathbf{T}_t^{\Sigma} \\ \hline \mathbf{T}^{\Sigma} & \mathbf{0} \end{array}\right] \tag{4.18}$$

$$\mathbf{T}^{\Sigma} = \mathbf{T}\mathbf{T}^{\mathrm{II}}\mathbf{T} = -je^{-j\varphi/2} \begin{bmatrix} \sin(\varphi/2) & \cos(\varphi/2) \\ \cos(\varphi/2) & -\sin(\varphi/2) \end{bmatrix}$$

It follows from Eqs. (4.18) that we can control the amount of power at the ports of the resultant composite directional coupler by varying the phase shift φ anywhere between zero and π.

Basically, the circuit in Fig. 4.19 operates as follows. The first constituent network splits the input signal applied to any port in group α into two equal parts. The phase relationship between these

two parts is adjusted in the second network. If phase-shift adjustment is accompanied by a marked attenuation in power, it will be preferable to use a separate phase shift in each channel, so that in each channel it can be changed by an equal amount but in opposite sense. The third constituent network operates as a power combiner: depending on the phase shift between the two equal signals at the ports on the left-hand side, it adds together the signals at one of the ports or subtracts one signal from the other at another port on the right-hand side. The circuit of Fig. 4.19 permits the use of any 3-dB directional couplers and any phase shifters and may be used as a prototype in the synthesis of a range of major microwave building blocks.

Notably, the inclusion of nonreciprocal phase shifters (they are non-adjustable phase shifters) in the second network leads to a four-port phase circulator. The use of a frequency-dependent reciprocal phase shifter (say, a transmission-line section of a length substantially longer than the wavelength in the guide) and broadband

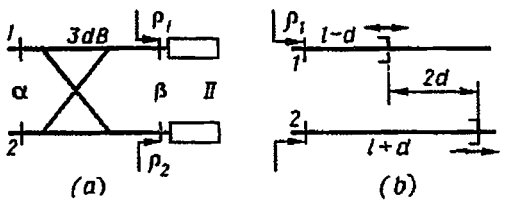


Fig. 4.20. Waveguide junction with the magnitude and phase of the reflection coefficient independently adjustable: *a*, general arrangement; *b*, loading stubs

directional couplers turns the circuit in Fig. 4.19 into a phase-type diplexer or channel-dropping filter. A mixture of two different frequencies ω_1 and ω_2, when applied to one of the ports on the left-hand side will be separated at the ports on the right-hand side so that the signal appearing at one of them will have frequency ω_1, and that appearing at the other port will have frequency ω_2. According to Eq. (4.18), the frequency-dependent phase-shifter of the filter will produce a phase shift equal to $\varphi_1 = 2n\pi$ ($n = 1, 2, 3, \ldots$) at frequency ω_1, and a phase shift equal to $\varphi_2 = \varphi_1 \pm \pi$ at frequency ω_2.

Note also that the function of the uncoupled transmission lines in the second network of the circuit in Fig. 4.19 can be performed by orthogonally polarized waves sustained in one and the same waveguide (for example, two modes orthogonally and linearly polarized in a circular or a rectangular waveguide). This offers an opportunity to use the circuit of Fig. 4.19 as a prototype for various polarizers constructed in circular and rectangular waveguide.

Example 4.7. A waveguide junction with the magnitude and phase of the reflection coefficient independently adjustable (Fig. 4.20*a*).

This junction contains a quadrature equal power divider (a quadrature 3-dB coupler) for which the scattering matrix is

$$\mathbf{S}^{\mathrm{I}} = \left[\begin{array}{c:c} \mathbf{0} & \mathbf{T}_t \\ \hdashline \mathbf{T} & \mathbf{0} \end{array}\right]$$

$$\mathbf{T} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -j \\ -j & 1 \end{bmatrix}$$

and which is terminated at ports β into two adjustable one-ports for which the diagonal scattering matrix has the form

$$\mathbf{S}^{\mathrm{II}}_{\beta\beta} = \{\rho\} = \begin{bmatrix} \dot{\rho}_1 & 0 \\ 0 & \dot{\rho}_2 \end{bmatrix}$$

Using the upper line in (4.15) for $\mathbf{S}^{\mathrm{I}} = 0$, we find the scattering matrix for the composite two-port made up of the 3-dB directional coupler and the terminations:

$$\mathbf{S}^{\Sigma}_{\alpha\alpha} = \mathbf{T}_t \{\rho\} \mathbf{T} = \frac{1}{2} \begin{bmatrix} (\dot{\rho}_1 - \dot{\rho}_2) & -j(\dot{\rho}_1 + \dot{\rho}_2) \\ -j(\dot{\rho}_1 + \dot{\rho}_2) & -(\dot{\rho}_1 - \dot{\rho}_2) \end{bmatrix}$$

Suppose that the terminations are shorted stubs adjustable in length (Fig. 4.20*b*) for which the reflection coefficients are

$$\dot{\rho}_1 = -\exp[-j2\beta(l-d)]$$

and

$$\dot{\rho}_2 = -\exp[-j2\beta(l+d)]$$

Then the scattering matrix of the composite two-port will have the form

$$\mathbf{S}^{\Sigma}_{\alpha\alpha} = j[\exp(-j2\beta l)] \begin{bmatrix} -\sin 2\beta d & \cos 2\beta d \\ \cos 2\beta d & \sin 2\beta d \end{bmatrix}$$

The two-port with the scattering matrix defined above offers an opportunity to adjust the magnitude of the reflection coefficient $s_{11} = |\sin 2\beta d|$ anywhere between 0 and 1 (this can be done by varyng the difference in length, $2d$, between the stubs while maintaining their average length l constant) and the phase of the reflection coefficient, $\arg \dot{s}_{11} = -(2\beta l + \pi/2)$, which can be done by varying the average length l of the stubs, while holding their difference in length, $2d$, at a constant value.

The circuit arranged as shown in Fig. 4.20*a* operates basically in about the same manner as the directional coupler discussed in

Example 4.6. The 3-dB directional coupler splits the input signal into equal parts, each of which undergoes a change in both magnitude and phase as they are reflected from the terminations. Then the same 3-dB directional coupler adds together or subtracts the reflected signals. If in the circuit of Fig. 4.20*a* the terminations are negative-resistance one-ports (say, tunnel diodes or reflecting parametric amplifiers), a microwave feed-through power amplifier will be formed. The amplified reflected signals will be combined at port *2* and cancel out at port *1*, and, importantly, these two ports can readily be interchanged.

In the circuit of Fig. 4.20*a* the quadrature 3-dB directional coupler may well be replaced by any other hybrid, say, a rat-race. We leave it as an exercise for the reader to set up the scattering matrix $S^{\Sigma}_{\alpha\alpha}$ applicable to such hybrids.

Chapter Five

Microwave Filters and Matching Circuits

5.1. Basic Definitions

A *microwave filter* is a passive two-port junction which serves to transmit microwave energy to a matched load so as to give a specified attenuation-vs-frequency response (attenuation characteristic).

The attenuation characteristic of a microwave filter is specified as follows. To begin with, one defines the pass band, that is, the frequency range within which the insertion loss, or attenuation (the reciprocal of the magnitude of the transmission coefficient expressed in decibels)

$$L\ (\text{dB}) = -20 \log |\, s_{21} \,|$$

does not exceed some specified value. Outside the pass band (that is, within the stop or rejection band), the insertion loss, or attenuation, must be as high as possible. Sometimes, one also qualifies the frequency behaviour of the phase of the transmission coefficient (usually as the limits of deviation from linearity over the bandwidth).

All filters may be classed into *low-pass filters*, *high-pass filters*, *bandpass filters*, and *band-stop* (or *band-rejection*) *filters*. The respective plots of attenuation-vs-frequency and microwave symbols of these filters are given in Fig. 5.1. The significant points on the plots are the *upper* and *lower cutoff frequencies* (also known as the *upper* and *lower band edge points*), ω_{co} (or ω_e) and the attenuation tolerance, L_1.

To minimize attenuation within the pass band, microwave filters are usually made with reactive (lossless and reciprocal) elements. An

abrupt increase in attenuation outside the pass band is produced by a nearly total reflection of the microwave energy from the filter input.

There are networks which are closely related to microwave filters in both structure and principle of design and known as *broadband matching circuits.* They, too, are reactive two-ports in the sense defined above and can serve either to provide the highest attainable TWR over the specified frequency band or to give a maximum band

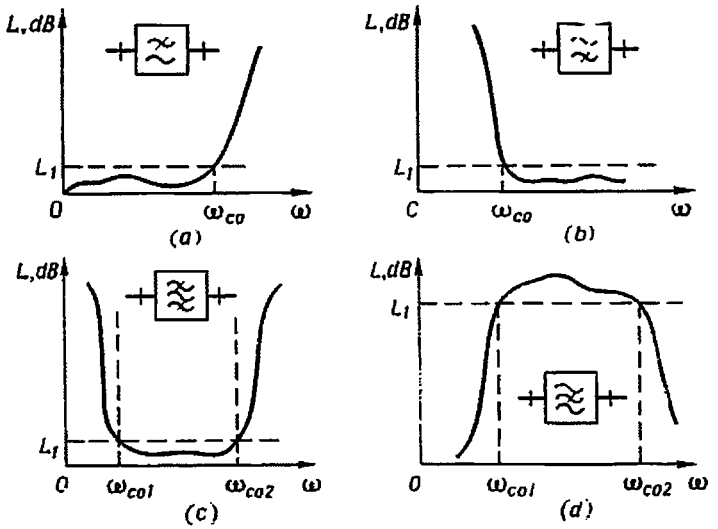


Fig. 5.1. Attenuation-vs-frequency characteristics of microwave filters: *a*, low-pass filter; *b*, high-pass filter; *c*, bandpass filter; *d*, band-stop (band-rejection) filter

of match at the specified TWR level. Some broadband matching circuits are intended for use with frequency-independent *resistive* loads, while others are tailored to operate with frequency-dependent impedance loads. In the latter case, the elements of a matching circuit are especially difficult to calculate.

In turn, two-port microwave filters with variously located pass and stop bands serve as building blocks for the more complex *frequency-division-multiplex* (FDM) microwave systems. Apart from their use in microwave transmission (feeder) systems, broadband matching circuits and filters are also used to provide interstage coupling in radio receivers and radio transmitters.

5.2. Filter Prototypes with an Optimal Attenuation-vs-Frequency Response

As a rule, a microwave filter is a cascade connection of several sections. Filter sections may be various resonators, stubs, coupled line segments, and so on. It is not at all easy to calculate the parameters of real filter sections from the specified frequency response of the filter. It appears more convenient to begin with some simplified equivalent circuit which may only approximately resemble the contemplated microwave filter. We will call this equivalent circuit a *prototype*. A prototype must permit an analytical calculation of the element values from the desired attenuation characteristic, $L(\Omega)$, where Ω is some normalized frequency related to the prototype. The most commonly used prototypes are lumped-LC four-terminal networks or two-ports.

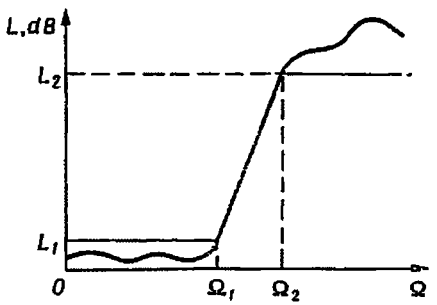


Fig. 5.2. Parameters determining the frequency response of a low-pass filter

Let us see how an LC-prototype can implement the desired frequency response of a low-pass filter. The requirements for the frequency response of this filter are explicitly stated in Fig. 5.2 which gives two attenuation levels, L_1 and L_2, and two band edge (or cutoff) frequencies, Ω_1 and Ω_2. The objective is to synthesize a prototype filter which will have a minimum count of circuit components and an insertion loss not over L_1 within the pass band (from 0 to Ω_1) and not less than L_2 in the stop band (from Ω_2 to infinity).

The first step in the synthesis procedure is to choose an analytical expression for the attenuation function. This expression must satisfy the condition of physical realizability; notably, it ought not to lead to a two-port containing negative inductances and capacitances. It can be shown that the physical realizability conditions will not be violated if we choose an attenuation function of the form

$$L(\Omega) = 1/|\dot{s}_{21}(\Omega)|^2 = 1 + [P_1^2(\Omega) + P_2^2(\Omega)]/Q^2(\Omega) \qquad (5.1)$$

where $P_1(x)$, $P_2(x)$ and $Q(x)$ are polynomials in the unknown variable x.

An ideal low pass-filter would have a zero attenuation within the pass band and an infinite attenuation outside the pass band. In other words, in an ideal low-pass filter the pass-band edge frequency Ω_1 and the stop-band edge frequency Ω_2 would be the same, and the attenuation characteristic would have a vertical portion. In a real

low-pass filter, however, the characteristic has an up-sloping shape with a slope decided by the degree of the polynomials in Eq. (5.1) which varies with the number of filter sections. Obviously, with a finite number of filter sections an ideal frequency response cannot be obtained. It can only be approximated in one of several ways. The approximation improves with increasing degree of the polynomials $P_{1,2}$ and Q.

Most often, filters are synthesized using the Butterworth (or maximally flat) approximation and the Chebyshev (or equal ripple) approximation.

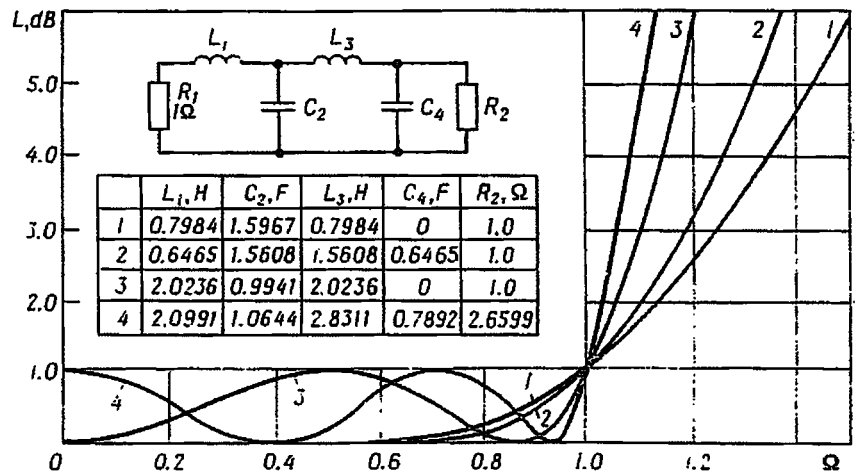
The Butterworth (maximally flat) attenuation characteristic. This form of the attenuation characteristic for a low-pass filter is written analytically as

$$L(\Omega) = 1 + \gamma^2\Omega^{2n}, \quad n = 1, 2, 3, \ldots \tag{5.2}$$

and is shown graphically in Fig. 5.3 for $n = 3$ and $n = 4$ (curves *1* and *2*, respectively). The attenuation level at the pass band edge, with $\Omega_{co} = 1$, is given by

$$L_1 = 10 \log (1 + \gamma^2)$$

The coefficient γ defines the attenuation tolerance within the pass band. At $\Omega > 1$, the function in (5.2) increases without bound as Ω



	L_1, H	C_2, F	L_3, H	C_4, F	R_2, Ω
1	0.7984	1.5967	0.7984	0	1.0
2	0.6465	1.5608	1.5608	0.6465	1.0
3	2.0236	0.9941	2.0236	0	1.0
4	2.0991	1.0644	2.8311	0.7892	2.6599

Fig. 5.3. Optimal frequency response of a ladder-type low-pass prototype filter

increases, and the rate of increase rises with increasing number n. When $\Omega < 1$, the function (5.2) runs progressively closer to the Ω-axis as n increases.

Maximally-flat (or Butterworth) filters are used in cases where high selectivity within the stop band is not required but there are stringent requirements as regards the quality of match within the pass band. An important advantage of the frequency response defined in (5.2) is that the phase dependence of the transmission coefficient shows a good linearity, a feature promoting an undistorted transmission of pulse waveforms.

The Chebyshev (equal-ripple) attenuation characteristic. In this case, the analytical expression for the attenuation characteristic of a low-pass filter has the form

$$L(\Omega) = 1 + \gamma^2 T_n^2(\Omega), \quad n = 1, 2, 3, \ldots \tag{5.3}$$

where γ is a real parameter defining the attenuation level

$$L_1 = 10 \log (1 + \gamma^2)$$

in the pass band, and $T_n(\Omega)$ is a Chebyshev polynomial* of the first kind, of degree n.

Let us recapitulate the basic properties of the Chebyshev polynomials. Low-degree Chebyshev polynomials of the first kind have the form

$$T_0(x) = 1$$

and

$$T_1(x) = x$$

The higher-degree polynomials are defined by a well-known recursive relation

$$T_{n+1} = 2xT_n - T_{n-1}$$

These polynomials show an oscillating behaviour in the interval $-1 \leqslant x \leqslant 1$ where their value varies between the limits ± 1. At $|x| > 1$, the Chebyshev polynomials abruptly increase in magnitude. The most important property of Chebyshev polynomials of the first kind is that in the interval $-1 \leqslant x \leqslant 1$ they are the least deviating ones from zero among all polynomials of degree n. This implies that the magnitude of any other polynomial of degree n with real coefficients and with the same coefficients of the highest term will exceed unity at some points on the interval $|x| \leqslant 1$. Basing ourselves on this property of the Chebyshev polynomials, we may argue that the equal-ripple attenuation characteristic defined in (5.3) gives the best approximation to the ideal square-shaped response at a fixed degree n, that is, for the specified number of filter elements. Examples of the Chebyshev attenuation characteristic for low-pass filters at $n = 3$ and $n = 4$ are shown in Fig. 5.3 (curves *3* and *4*, respectively).

* The symbol "T" for Chebyshev polynomials is used because the name of the scientist is often written as "Tchebyshev". In fact, many publications use the symbol "C" for these polynomials.—*Translator's note.*

The exponent n, which defines the number of elements in the prototype filter, can be found from the requirements to be met by the attenuation characteristic of the filter (see Fig. 5.2). On denoting

$$L_1 = 10 \log L_p$$

and

$$L_2 = \log L_s$$

where L_p and L_s are the required attenuation level at the pass band edge frequency Ω_1 and at the stop band edge frequency Ω_2, then for a maximally flat (Butterworth) filter

$$n \geqslant \frac{\log [(L_s - 1)/(L_p - 1)]^{1/2}}{\log (\Omega_2/\Omega_1)}$$

and for an equal-ripple (Chebyshev) filter

$$n \geqslant \frac{\operatorname{arcosh} [(L_s - 1)/(L_p - 1)]^{1/2}}{\operatorname{arcosh} (\Omega_2/\Omega_1)}$$

From comparison of the two expressions for n we can see that, given the same requirements for the attenuation characteristic, the Chebyshev filter requires fewer elements.

Both the Butterworth and the Chebyshev attenuation characteristics are implemented by what is called the *ladder prototype circuit* (Fig. 5.3). Given the responses of the form defined in (5.2) or (5.3), it would require to use a rather unwieldily algorithm for the computation of the L_i and C_i element values, so in practice it is customary to employ ready-made computer programs or to look up reference tables. The terminations in the prototype of Fig. 5.3 have a value of unity for any n in the case of the Butterworth filter and for an odd n in the case of the Chebyshev filter. If we take the Chebyshev attenuation characteristic and even values of n, then the attenuation at zero frequency must be equal to L_1, and the required reflection coefficient $| s_{11} | = [(L_p - 1)/L_p]^{1/2}$ can only be obtained with one of the terminations chosen to be other than unity.

5.3. Change of the Frequency Variable in the Design of Filters

The tabulated values of L_i and C_i for ladder-type low-pass filters are normalized with respect to a band edge frequency of unity, $\Omega_e = 1$. As an illustration, Fig. 5.3 shows the element values related to the corresponding attenuation plots and taken from reference tables. If we want to make calculations at other band-edge values or for other types of filter (high-pass, band-pass, or band-stop), we must *change the frequency variable* Ω in some definite manner. In doing so, it is important to remember that the attenuation functions

of two-ports are even with respect to frequency, that is,

$$L(\omega) = L(-\omega)$$

To begin with, let us consider the change of the frequency variable equivalent to a change of frequency scale:

$$\Omega = K_1\omega$$

where Ω is the normalized band-edge frequency, K_1 is a real constant coefficient, and ω is the actual frequency.

In cases involving the lumped L and C circuit parameters or constants (and also the mutual inductance M), frequency turns up in all equations for the circuit parameters as the factor of L, C, and M. If we choose

$$K_1 = 1/\omega_e$$

where ω_e is the actual band-edge frequency, we can always make the normalized band-edge frequency of the circuit in question equal to unity, $\Omega_e = 1$. Then the reactances C, L and M must be multiplied by the actual band-edge frequency. If, on the other hand, we have obtained a normalized low-pass prototype with a band edge frequency of unity, this prototype may be used for conversion to any actual pass band edge frequency ω_e by dividing the values of all the reactances in the prototype by this frequency. That is the rationale of frequency normalization.

A second change of the frequency variable

$$\Omega = -K_2/\omega \tag{5.4}$$

is equivalent to interchanging the origin of coordinates and a frequency point at infinity, and also the positive and negative directions of the frequency axis. As a result, the attenuation characteristic of a low-pass filter is transformed into that of a high-pass filter, as shown in Fig. 5.4*a*. The change of the frequency variable as defined in (5.4) transforms the filter reactances as follows:

$$\Omega L \rightarrow -K_2L/\omega = -1/\omega C'$$
$$\Omega C \rightarrow -K_2C/\omega = -1/\omega L'$$

where $K_2 = \omega_e$, if $\Omega_e = 1$. Therefore, the circuit of the respective high-pass filter (Fig. 5.4*b*) is derived from the low-pass prototype with $\Omega_e = 1$ through the replacement of each inductance L by a capacitance such that

$$C' = 1/\omega_e L$$

and of each capacitance C by an inductance such that

$$L' = 1/\omega_e C$$

A third change of the frequency variable takes the form

$$\Omega = K_3\omega_0\,(\omega/\omega_0 - \omega_0/\omega) \tag{5.5}$$

where K_3 and ω_0 are positive constants.

According to (5.5), the even function $L\,(\Omega)$ transforms into the even function $L\,(\omega)$ which shows symmetry with respect to points ω_0 and $-\omega_0$ on the positive and negative directions of the frequency

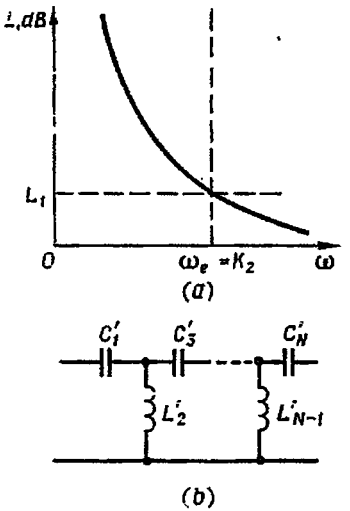


Fig. 5.4. High-pass filter: (*a*) frequency response and (*b*) circuit after conversion from a low-pass prototype

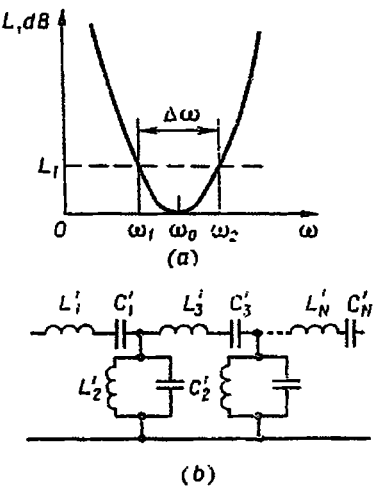


Fig. 5.5. Band-pass filter: (*a*) frequency response and (*b*) circuit after conversion from a low-pass prototype

axis. If we choose $L\,(\Omega)$ to be the attenuation characteristic of a low-pass filter (see Fig. 5.3), then the transformation (or mapping) (5.5) will lead to the attenuation characteristic of a band-pass filter (Fig. 5.5*a*). The upper and lower band edge points, ω_1 and ω_2, of a band-pass filter are connected to the band-edge of the prototype low-pass filter, $\Omega_e = 1$, by relations of the form

$$\omega_2 - \omega_1 = \Delta\omega = 1/K_3$$

and

$$\omega_1\omega_2 = \omega_0^2$$

When the frequency variable is changed as defined in (5.5), any inductance L in the prototype low-pass filter with a band edge frequency of unity becomes a series resonant circuit formed by an inductance

$$L' = L/\Delta\omega$$

and a capacitance

$$C' = \Delta\omega/\omega_0^2 L$$

At the same time, any capacitance in the low-pass prototype becomes a parallel resonant circuit formed by a capacitance

$$C'' = C/\Delta\omega$$

and an inductance

$$L'' = \Delta\omega/\omega_0^2 C$$

The changes that take place in the prototype circuit as a result of the low-pass to band-pass mapping according to Eq. (5.5) are shown in Fig. 5.5*b*.

If we apply the frequency scaling defined in (5.5) to a high-pass filter, we shall obtain a band-stop filter whose attenuation chara-

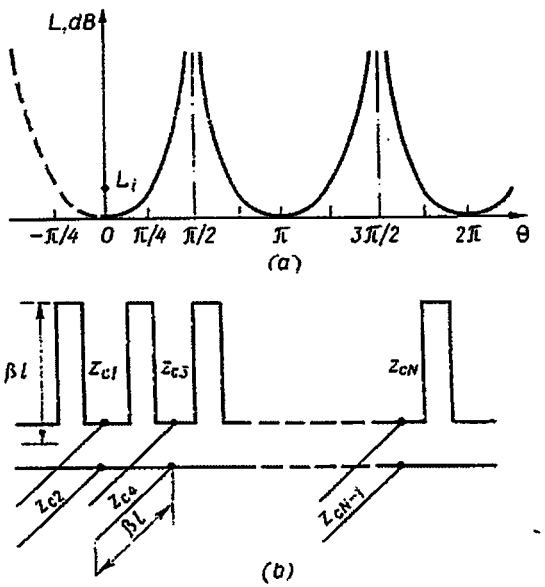


Fig. 5.6. Equal-length-section filter: (*a*) attenuation-vs-frequency characteristic and (*b*) circuit after conversion from a low-pass prototype

cteristic is symmetrical about the central frequency ω_0. Hence, in order to derive a band-stop filter from a low-pass prototype, we must consecutively apply two frequency transformations, one as defined in (5.4) and the other as defined in (5.5).

To sum up, any of the frequency-selective filters examined above can be derived from an only prototype—the low-pass filter with a band edge frequency of unity.

One more change of frequency variable is worth mentioning:

$$\Omega = \tan \beta l \tag{5.6}$$

where βl is the electrical length of a transmission-line section.

As follows from Eq. (5.6), the even attenuation function $L(\Omega)$ associated with a lumped-LC prototype filter transforms into a periodic function, $L(\beta l)$. For example, the attenuation characteristic of the low-pass prototype shown in Fig. 5.3 takes the shape shown in Fig. 5.6*a*. The reactance of any inductance in the low-pass prototype becomes the reactance of a shorted stub,

$$\Omega L_i \rightarrow z_{ci} \tan \beta l$$

such that the normalized characteristic impedance of the stub is numerically equal to the inductance (Fig. 5.6*b*). Similarly, the reactance of any capacitance becomes the susceptance of an open-circuited stub,

$$\Omega C_i \rightarrow (1/z_{ci}) \tan \beta l$$

such that the characteristic impedance of the stub is numerically equal to the reciprocal of the capacitance. The transformation defined by Eq. (5.6) is due to Richards. It offers an opportunity to reduce the analysis and synthesis of microwave networks fabricated in equal-length transmission-line sections in any combinations with resistors to that of lumped-constant circuits.

5.4. Use of Transmission-Line Segments in Microwave Filters

After the designer has chosen an optimal prototype filter and changed the frequency variable, he faces one more crucial question: How can the ideal lumped-constant elements of the prototype be replaced at microwave frequencies? The problem may be tackled in more than one way, and the answer depends on the frequency band in which the contemplated filter is to operate, the type of transmission line used, and the relative bandwidth of the filter.

Most often, lumped capacitances, inductances, and resonant circuits are replaced by transmission-line segments. This approach is especially convenient if the relative bandwidth of a filter is in excess of 5% of the centre frequency. The replacement of lumped inductances and capacitances by shorted or open-circuited transmission-line segments of reduced length (less than a quarter of the wavelength in the guide) is straightforward. Examples of such filter sections constructed in coaxial line and the applicable equivalent circuits are shown in Fig. 5.7.

The series and parallel resonant circuits of band-pass filters can conveniently be fabricated in resonant transmission-line segments.

Shorted segments are used most often (Fig. 5.8). The series and parallel resonant circuits of filters should be replaced by resonant stubs in such a way that the resonance frequencies of the filter and the external Qs of the resonant circuits remain unchanged.

As will be recalled, the Q-factor of any resonant circuit is the ratio of the average stored electromagnetic energy to the energy dis-

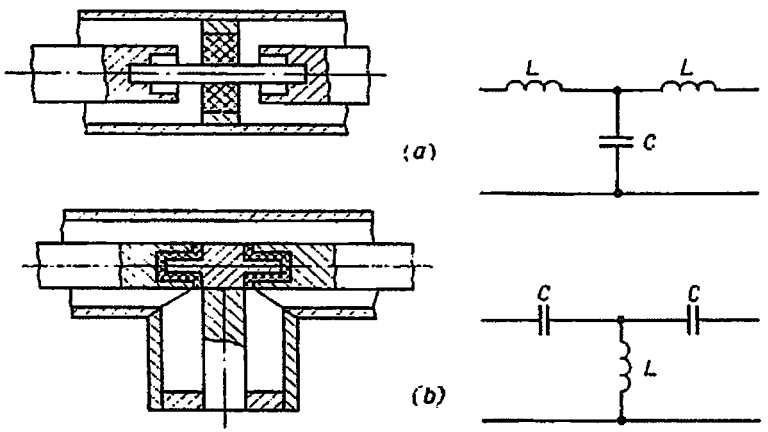


Fig. 5.7. LC-sections using short transmission lines for (a) a low-pass filter and (b) a high-pass filter and the respective equivalent circuits

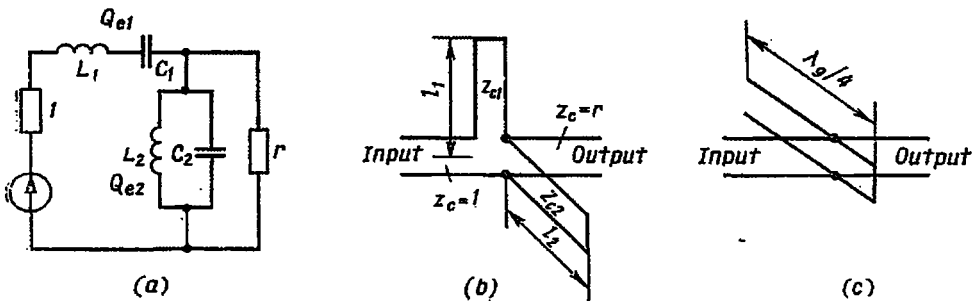


Fig. 5.8. Use of resonant transmission-line sections in a bandpass filter: a, prototype; b, realization in shorted stubs; c, stub tapped down on a transmission line

sipated over a cycle at the resonant frequency ω_0. For filters, a more convenient equation is

$$Q = \frac{\omega_0}{2r_0}\left|\frac{dx(\omega)}{d\omega}\right|_{\omega=\omega_0} = \frac{\omega_0}{2g_0}\left|\frac{db(\omega)}{d\omega}\right|_{\omega=\omega_0} \qquad (5.7)$$

because it is equally applicable to series and parallel resonance. Here, $x(\omega)$ and $b(\omega)$ are the reactance and susceptance of the reso-

nant circuit in question, while r_0 and g_0 are the series loss resistance and the shunt loss conductance at the resonance frequency ω_0.

When determining the *external* Q, or Q_e, of a resonant circuit, it is assumed that the input power is solely dissipated in the external resistive loads connected to the resonant circuit, whereas the internal ohmic losses are neglected. In contrast, when determining the internal or *unloaded* Q of a resonant circuit, Q_u, the power drawn by the external resistive loads is ignored and account is taken of only the ohmic losses caused by the departure of the conductors and dielectrics in the resonant circuit from an ideal behaviour. Lastly, when determining the *loaded* Q of a resonant circuit, Q_l, one takes into account the losses both inside and outside the resonant circuit. Accordingly, the loaded, external and unloaded Q-factors of a resonant circuit are connected by a relation of the form

$$1/Q_l = 1/Q_e + 1/Q_u$$

Lossless filters are built such that $Q_u \gg Q_e$, so we may, in a first approximation, neglect the internal losses in the resonant circuits.

In accordance with (5.7), the external Q of a series resonant circuit (Fig. 5.8*a*) is given by

$$Q_{e1} = \frac{\omega_0}{2(1+r)} \frac{\mathrm{d}}{\mathrm{d}\omega} (\omega L_1 - 1/\omega C_1)_{\omega=\omega_0} = \omega_0 L_1/(1+r)$$

where

$$\omega_0 = 1/(L_1 C_1)^{1/2}$$

is the resonance frequency, and $r + 1$ is the total external loss resistance equal to the sum of the left-hand and right-hand load resistances of the filter.

The external Q of a parallel resonant circuit (Fig. 5.8*a*) is

$$Q_{e2} = \frac{\omega_0}{2(1+1/r)} \frac{\mathrm{d}}{\mathrm{d}\omega} (\omega C_2 - 1/\omega L_2)_{\omega=\omega_0} = \omega_0 C_2 r/(1+r)$$

where

$$\omega_0 = 1/(L_2 C_2)^{1/2}$$

and $1 + 1/r$ is the sum of the conductances of the left- and right-hand loads of the filter.

The above equations offer an opportunity to find the required value of the external Q for the resonant circuits of a filter from the element values of the low-pass filter prototype.

When determining the external Q of a resonant stub (or of any other implementation of a resonant circuit), the first thing to do is to write expressions for the input impedance or admittance of each stub with allowance for the terminations, then to use Eqs. (5.7). The input impedance of the series stub (Fig. 5.8*b*), including the

two terminations, is

$$\dot{z}_1 = (1 + r) + jz_{c1} \tan (2\pi l_1/\lambda_g)$$

At points of series resonance,

$$l_1 = k\lambda_g/2$$

Therefore, the external Q of the series stub is

$$Q_1 = \frac{\omega_0}{2(1+r)} \frac{dx_1}{d\omega}\bigg|_{\omega=\omega_0} = \frac{k\pi z_{c1}}{2(1+r)}, \quad k = 1, 2, 3, \ldots$$

The input admittance of the shunt stub, including those of the terminations, may be written as

$$\dot{y}_2 = (r + 1)/r - j\,(1/z_{c2}) \cot (2\pi l_2/\lambda_g)$$

and the external Q at points of parallel resonance for $l_2 = (2m - 1) \times \lambda_g/4$ can be found by the equation

$$Q_2 = \frac{\omega_0 r}{2(r+1)} \frac{db_2}{d\omega}\bigg|_{\omega=\omega_0} = \frac{\pi r (2m-1)}{4(r+1) z_{c2}}, \quad m = 1, 2, 3, \ldots$$

The equations for Q_1 and Q_2 enable the wave impedances of the series and shunt stubs to be determined from the required values of the external Qs in the prototype filter, whereas the stub lengths should be chosen so as to retain the required resonant frequency. When resonant LC-circuits are replaced by resonant stubs, it may so happen that the resultant characteristic impedances of the stubs are unacceptable from engineering considerations. This difficulty can be circumvented by choosing appropriate values for k and m, and also by tapping the stubs down on the transmission line (Fig. 5.8*c*).

Obviously, resonant stubs may be treated as a kind of cavity resonators (or resonant cavities). Since the ratio of volume to surface area in such resonators is small, their unloaded Q does not usually exceed 500. This is the reason why when filters are constructed in resonant stubs, the external Qs of the individual stubs ought not to exceed 100 to 150.

5.5. Resonators in the form of Discontinuities in a Transmission Line

In band-pass microwave filters use is often made of non-reflecting resonators formed by a pair of spaced discontinuities on a transmission line or a waveguide. Figure 5.9*a* shows such a resonator formed by two identical discontinuities shunting a transmission line. The discontinuities may be, say, waveguide diaphragms (irises) or other reactances representable by a similar equivalent circuit. If the susceptance $\tilde{b}$ tends to infinity (that is, if the physical opening in the diaphragm progressively decreases in size), then in the limit a transmission-line section will be formed, shorted at both ends and

acting as a resonator. At high, but finite values of susceptance $\tilde{b}$ there appears an opportunity for an exchange of energy between the resonator and the incoming transmission lines. The amount of coupling between the resonator and the lines can be adjusted along with its bandwidth by varying the value of the susceptance.

The resonator in Fig. 5.9*a* consists of three cascaded elementary two-ports, and its classical cascading (ABCD) matrix may be written as the product of three elementary cascading matrices. Neglecting losses, the cascading matrix of the resonator may be put in the following form:

$$\begin{bmatrix} \dot{a} & \dot{b} \\ \dot{c} & \dot{d} \end{bmatrix} = \begin{bmatrix} \cos\beta l - \tilde{b}\sin\beta l & j\sin\beta l \\ j[2\tilde{b}\cos\beta l + (1-\tilde{b}^2)\sin\beta l] & \cos\beta l - \tilde{b}\sin\beta l \end{bmatrix} \quad (5.8)$$

where the susceptance $\tilde{b}$ and the electrical length βl between the discontinuities are, in the general case, frequency-dependent.

Let us require at the resonance frequency for the input reflection coefficient of the resonator to vanish and, as a consequence, for all of the power applied to the input to be transmitted to the output. According to the conversion equation (4.5), this condition takes the form

$$(\dot{a} - \dot{d}) + (\dot{b} - \dot{c}) = 0$$

Owing to the symmetry of the resonator, $\dot{a} = \dot{d}$, and so the condition of match reduces to a simple equality, $\dot{b} = \dot{c}$. Equating the cascading coefficients $\dot{b}$ and $\dot{c}$ in (5.8) leads us to the following equation:

$$\sin\beta l = 2\tilde{b}\cos\beta l + (1-\tilde{b}^2)\sin\beta l$$

which may be treated as the *condition of resonance*. The roots of this quadratic equation in $\tilde{b}$ are

$$\tilde{b}_0' = 2\cot\beta_0 l$$

and

$$\tilde{b}_0'' = 0$$

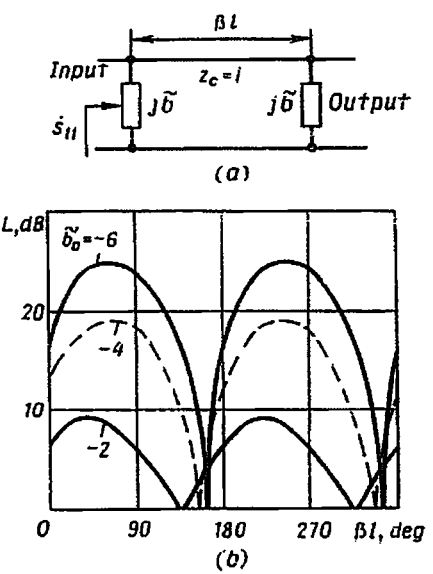


Fig. 5.9. Non-reflecting shunt-discontinuity resonator: (*a*) circuit and (*b*) attenuation-vs-frequency characteristic

where the second root, $\tilde{b}_0'' = 0$, represents the trivial case of a uniform transmission line and may therefore be discarded. Thus, the

transcendental equation giving resonance frequencies may be written as

$$\tilde{b}_0(\omega_0) = 2\cot(\omega_0 l/v_p) \tag{5.9}$$

where v_p is the phase velocity.

The susceptance of a discontinuity is a slowly varying function of frequency, and this offers an opportunity to find from the known value of $\tilde{b}_0$ the resonator length l giving the requisite resonance frequency. Off resonance, the resonator reflection coefficient is nonzero and, in consequence, the transmission coefficient decreases. By definition, the attenuation characteristic of a resonator over the bandwidth has the form

$$L = 1/|\dot{s}_{21}(\omega)|^2 = |\dot{a} + \dot{b} + \dot{c} + \dot{d}|^2/4$$

which takes care of the connection between the elements of the scattering matrix and the elements of the transfer matrix. On substituting here the elements of the A matrix from (5.8) and collecting the like terms, we obtain an explicit expression for the attenuation function

$$L = 1 + \tilde{b}^2[\cos\beta l - (\tilde{b}\sin\beta l)/2]^2$$

It is an easy matter to verify that for the values of $\tilde{b}$ satisfying Eq. (5.9), the attenuation function reduces to unity. Typical plots of attenuation vs electrical length for various inductive discontinuities are shown in Fig. 5.9*b*. As a rule, this form of resonator is arranged to operate at $\beta l \approx \pi$ (a half-wave resonator).

In the vicinity of the resonant frequencies defined by (5.9), the attenuation function can be approximated by two terms from a Taylor series in the variable $\Phi = (\beta - \beta_0)\, l$:

$$L \approx 1 + \tilde{b}_0^2[1 + (\tilde{b}_0^2/4)]\,\Phi^2$$

where

$$\tilde{b}_0 = 2\cot\beta_0 l$$

Using an obvious definition for electrical length

$$\beta l = \omega l/v_p = (\omega l/c)(\mu_r\varepsilon_r)^{1/2}[1 - (\omega_c/\omega)^2]^{1/2}$$

where c is the velocity of light, μ_r is the relative permeability and ε_r is the relative permittivity of the dielectric filling the cross-section of a TE, TM or TEM transmission line, and ω_c is the corresponding critical frequency, we obtain on carrying out identity transformations

$$\Phi = \frac{\beta_0 l}{1-(\omega_c/\omega_0)^2}\,\frac{\Delta\omega}{\omega_0}$$

where

$$\Delta\omega = \omega - \omega_0$$

On substituting the above expression in the equation for L, we may write the equation for the insertion loss (attenuation) of the resonator in the vicinity of the resonance frequency in the canonical form

$$L \approx 1 + 4Q_e^2 (\Delta\omega/\omega_0)^2$$

where

$$Q_e = (\tilde{b}_0/2)\,(1 + \tilde{b}_0^2/4)^{1/2}\,\{\beta_0 l/[1 - (\omega_c/\omega_0)^2]\} \qquad (5.10)$$

has the meaning of the *external* Q-factor of the resonator.

If we take into account the ohmic losses as well, the resonator can be additionally characterized in terms of its unloaded Q-factor, Q_u, obtained when the lines are shorted, and in terms of its loaded Q-factor, Q_l, obtained from the equation

$$1/Q_l = 1/Q_e + 1/Q_u$$

If $Q_e \ll Q_u$, then the loaded Q is in effect the same as the external Q. In such a case, Q_e can be measured from the width of the resonance curve

$$Q_e = \omega_0/2\Delta\omega_{3dB}$$

where the bandwidth is taken between half-power (3-dB) points and the power transmission coefficient at the band edges is 0.5.

5.6. Microwave Filters Using Quarter-Wave and Direct-Coupled Resonators

In the ladder prototype of a band-pass filter, the alternating series and parallel resonant circuits (see Fig. 5.5*b*) are directly joined to one another, and this is a major inconvenience at microwave frequencies. This drawback is avoided in a quarter-wave prototype where the resonant circuits are connected in the transmission line so that they are spaced $\lambda_g/4$ apart.

The synthesis of quarter-wave-coupled filters is based on the equivalence of two 2-ports one of which is in the form of a normalized lumped impedance $\dot{z}$ inserted in series with the transmission line, and the other is in the form of a half-wave line section with a normalized lumped admittance $\dot{y} = \dot{z}$ which shunts the line at its middle point (Fig. 5.10). The equivalence is easy to establish by comparing the respective classical cascading matrices. Let us model the half-wave line segments shunted by the admittance $\dot{y}$ as a cascade connection of three elementary two-ports, namely: a uniform transmission line of length $\lambda_g/4$, a shunt admittance $\dot{y}$, and a second transmission

line likewise quarter-wave long. The classical cascading (ABCD) matrix for this cascade connection will have the form

$$\begin{bmatrix} \dot{a} & \dot{b} \\ \dot{c} & \dot{d} \end{bmatrix} = \begin{bmatrix} 0 & j \\ j & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \dot{y} & 1 \end{bmatrix} \begin{bmatrix} 0 & j \\ j & 0 \end{bmatrix} = \begin{bmatrix} -1 & -\dot{y} \\ 0 & -1 \end{bmatrix} = - \begin{bmatrix} 1 & \dot{z} \\ 0 & 1 \end{bmatrix}$$

As is seen, the classical cascading matrices of both two-ports are identical to within a sign. The difference in sign implies that the output voltages and currents in the second 2-port undergo one more phase reversal.

When building a quarter-wave-coupled filter, all series resonant circuits of the *LC*-prototype are replaced by lines half-wave long,

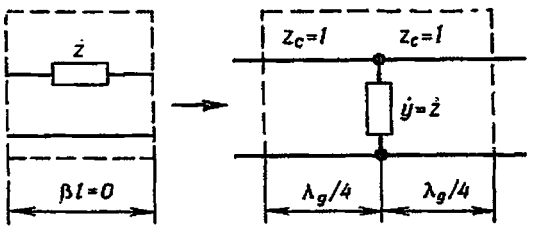


Fig. 5.10. Illustrating the principle of quarter-wavelength coupling

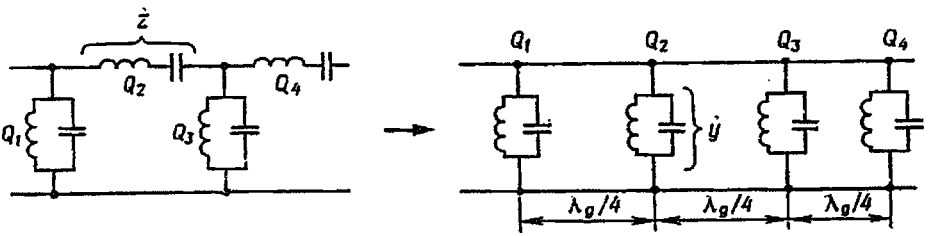


Fig. 5.11. Quarter-wavelength coupling in a band-pass prototype

shunted at the middle by parallel resonant circuits so that the filter circuit has the configuration shown in Fig. 5.11. The equivalence between the synthesized and prototype filters is complete only at the centre frequency where the transmission lines have a length exactly equal to a quarter of the wavelength in guide.

Instead of decoupling quarter-wave sections, use may be made of other two-ports called *invertors*. They have frequency-invariant cascading matrices identical with that of a quarter-wave transformer. If the relative bandwidth of the filter is narrow (of the order of 5 to 7% of the centre frequency), the frequency dependence of the connecting lines may be neglected. In filters with a broader bandwidth, the external *Q*s of the resonant circuits can be compensated

so as to make up for the dependence of the electrical length of the coupling sections on frequency.

In quarter-wave-coupled filters, the resonant circuits may be shunt stubs, resonant diaphragms, and various resonant cavities. By far the best choice seems to be the transmission resonator formed by two spaced discontinuities, discussed in Sec. 5.5. The use of such resonators in three-section quarter-wave-coupled filters is illustrated in Fig. 5.12. The filter uses inductive discontinuities

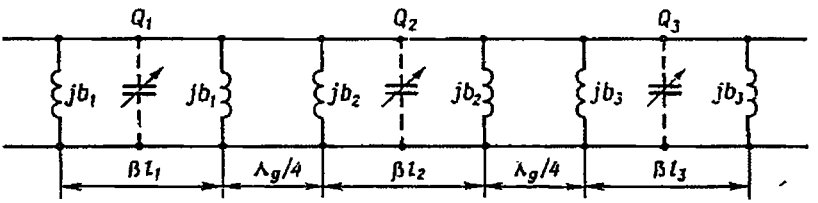


Fig. 5.12. Three-resonator, quarter-wave-coupled filter

which have, as a rule, a better electrical strength as compared with capacitive discontinuities. The susceptance of a discontinuity, required for each resonator, can be found from Eq. (5.10), proceeding from the known external Q of the lumped-parameter prototype fil-

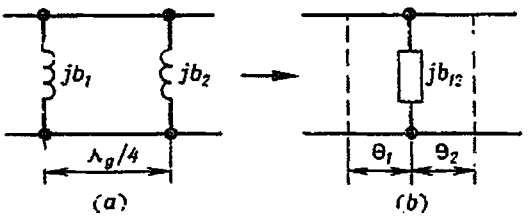


Fig. 5.13. Direct coupling between resonators

ter. The resonators can be tuned to the desired operating frequency by adjustable capacitive elements located in the middle of each resonator.

A major limitation of quarter-wave-coupled filters is an increased size because of the need to use transmission lines to interconnect the neighbouring resonators. This drawback can be avoided by directly coupling the resonators.

Directly coupled filters. Let us select in the quarter-wave-coupled filter of Fig. 5.12 two adjacent discontinuities, say, b_1 and b_2 along with the quarter-wave coupling line (Fig. 5.13*a*). These elements form a non-resonant reactive two-port for which the classical casca-

ding (ABCD) matrix is

$$\mathbf{A}_1 = \begin{bmatrix} 1 & 0 \\ jb_1 & 1 \end{bmatrix} \begin{bmatrix} 0 & j \\ j & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ jb_2 & 1 \end{bmatrix} = \begin{bmatrix} -b_2 & j \\ j(1-b_1b_2) & -b_1 \end{bmatrix}$$

Let us try to obtain a similar matrix, using an equivalent two-port which contains only one discontinuity, b_{12}, located between two short lines of electrical lengths $\theta_1 = \beta l_1$ and $\theta_2 = \beta l_2$, as shown in Fig. 5.13*b*. The cascading matrix for this equivalent two-port will have the form

$$\mathbf{A}_2 = \begin{bmatrix} \cos(\theta_1+\theta_2) - b_{12}\sin\theta_1\cos\theta_2 & j[\sin(\theta_1+\theta_2) - b_{12}\sin\theta_1\sin\theta_2] \\ j[\sin(\theta_1+\theta_2) + b_{12}\cos\theta_1\cos\theta_2] & \cos(\theta_1+\theta_2) - b_{12}\cos\theta_1\sin\theta_2 \end{bmatrix}$$

Equating the corresponding elements in the matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ and solving the resultant set of equations, we obtain the condition of equivalence between the coupling two-ports in Fig. 5.13*a* and *b*:

$$\tan(\theta_1 - \theta_2) = (b_1 - b_2)/b_1b_2, \; b_{12} = -b_1b_2/\cos(\theta_1 - \theta_2)$$

$$\sin(\theta_1+\theta_2) = \frac{(b_1+b_2)\,b_{12} + 2\,(2-b_1b_2)}{b_{12}^2+4}$$

The replacement of the coupling two-ports in the filter of Fig. 5.12 by equivalent circuits leads to a direct-coupled filter, such as shown

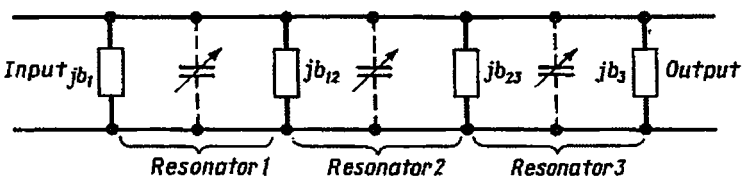


Fig. 5.14. Direct-coupled three-resonator filter

in Fig. 5.14. Direct coupling cuts down the size of the filter by 25 to 40% with a relative bandwidth of 1 to 10% of the centre frequency.

Dielectric-resonator filters. Cavity-resonator waveguide filters are highly stable in performance and suffer low losses, but they are too large in size and heavy in weight. Advances in the manufacture of high-quality dielectrics, such as ceramics based on $Ba_2Ti_9O_{20}$, $(Zr, Sn)\,TiO_4$ etc., have offered an opportunity in the last decade to build and use small-size *dielectric-resonator filters* in microwave transmission systems. They combine insignificant losses, small size, and low cost.

For its operation, a dielectric resonator depends on the resonance of electromagnetic waves within the bulk of a dielectric material. At resonance, electromagnetic energy is stored due to the repeated

superposition of practically in-phase waves travelling in opposite directions and reflected at the boundaries of the specimen. In metal cavity resonators, the waves are reflected from the metal walls; in dielectric resonators they are reflected from the interface between the dielectric and air. If the dielectric has a sufficiently high permittivity, the coefficient of reflection from the interface (especially from a curved one with a small radius of curvature) may be equal to unity very nearly — a fact which assures substantially smaller

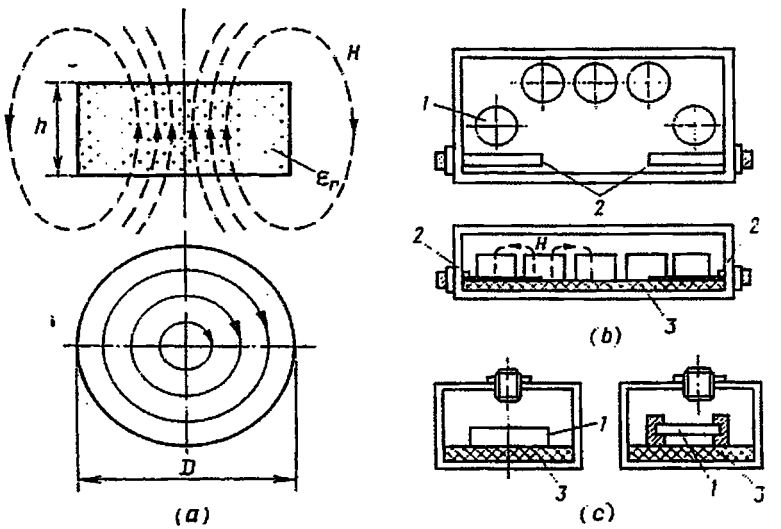


Fig. 5.15. Filter using dielectric disc resonators: *a*, field pattern of the $TE_{01\delta}$ mode due to an isolated resonator; *b*, direct-coupled, strip-line filter (*1*, resonator; *2*, strip conductor; *3*, substrate); *c*, placement of resonators and tuning screws

losses than in the case of reflections from true metals. The unloaded Q-factor, Q_u, of a dielectric resonator is practically decided solely by the losses in the dielectric, and for state-of-the-art materials with a relative permittivity of around 10 to 40 it may be anywhere from 5×10^3 to 1×10^4 in the range from decimetric (UHF) to short millimetric (EHF) waves. Apart from the unloaded Q, a parameter of importance for dielectric resonators is the temperature stability of the resonant frequency. For existing materials used in dielectric resonators it may be comparable with that of cavity resonators built of Invar.

Most often, small-size microwave filters use cylindrical dielectric resonators with a height-to-diameter ratio of anywhere from 0.3 to 0.5. The preferable $TE_{01\delta}$ mode in such a resonator produces the field pattern shown in Fig. 5.15*a*. For a resonator 6 mm in diameter

and with a relative permittivity of 36, the resonance frequency is around 8.5 GHz.

In multi-resonator filters it is convenient to employ directly coupled dielectric resonators. An example of a five-resonator bandpass filter constructed in strip line is shown in Fig. 5.15*b*. To prevent spurious radiation, the filter is enclosed in a metal shield which is in effect a below-cutoff rectangular waveguide.

The amount of coupling between the individual dielectric resonators required to give the desired bandwidth with a given frequency response (maximally flat or equal-ripple) is obtained by varying the relative position of the resonators on a strip-line board. The chosen resonant frequency can be trimmed to within $\pm 0.3\%$ of the optimal value with the aid of tuning screws arranged on the top wall of the shield coaxially with the resonator (Fig. 5.15*c*).

At present, dielectric-resonator filters come in a variety of designs, and intensive research is under way with a view to developing adequate procedures for the computer-aided design of filters.

5.7. Broadband Matching of Resistive Loads by Stepped-Impedance Transformers

The need to match resistive loads may arise, for example, when two transmission lines differing in cross-sectional dimensions or geometry are to be joined together. In the specifications for such joints it is usual to state some bandwidth and the mismatch tolerance over the bandwidth. Thus, the problem facing the designer is akin to that of building frequency-selective filters.

The lines being joined are usually matched by means of an intervening nonuniform line section called an *impedance matching transformer*. There may be *tapered-impedance transformers* if their cross-sectional dimensions change gradually and continuously along the length, and *stepped-impedance transformers* if their cross-sectional dimensions change abruptly, or stepwise, from section to section. If no constraints are imposed on the length of a transformer, the mismatch may be reduced to any value, however small, over any bandwidth. In practice, however, it is usually sought to make a transformer as short as practicable for a given change in characteristic impedance, the desired bandwidth, and the specified mismatch tolerance.

The simplest stepped-impedance transformer is the quarter-wave transformer; unfortunately, it has a limited bandwidth. An improvement in bandwidth along with an improvement in the quality of match can be obtained with multistep transformers. The step length, or the length of uniform portions of a transformer, is chosen to be the same for all steps, and the desired shape of the attenuation characteristic is secured through an appropriate choice of characteristic impedance for the steps.

Most often, resort is made to transformers having the *Chebyshev (equal-ripple)* or the *Butterworth (maximally-flat)* reflection-vs-frequency characteristic. The former are designed with the aid of Chebyshev polynomials and strike an optimum balance between the match bandwidth, the mismatch tolerance, and the transformer length. The reflection coefficient of Butterworth transformers remains constant with frequency over the match bandwidth, and their transmission coefficients behave almost linearly with both phase and frequency.

As an example, we shall go through the design procedure for a *Chebyshev three-step transformer*. Any one of two methods may be

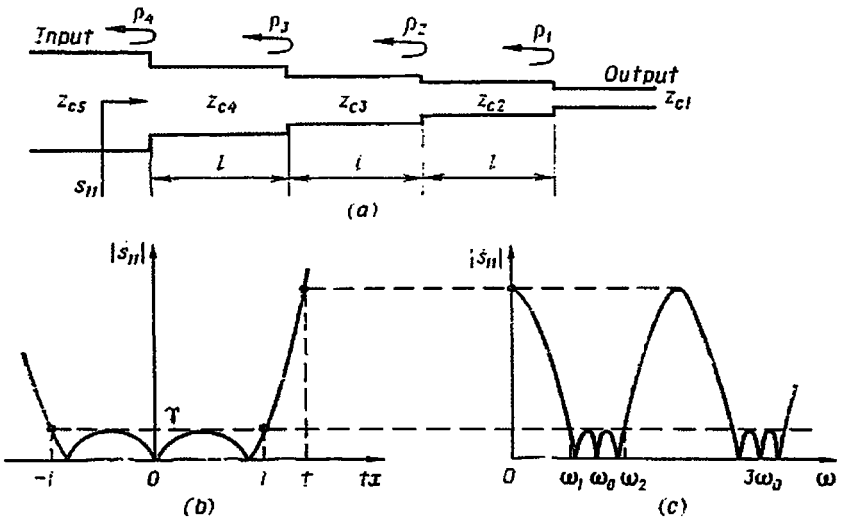


Fig. 5.16. Three-step (three-section) Chebyshev transformer: *a*, general arrangement; *b* and *c*, attenuation characteristics

used: an approximate method based on indeterminate coefficients, applicable when the number of steps is small, and a more rigorous method which, although more tedious, yields better results for any number of steps. We shall use the former method to obtain the reflection characteristic of the transformer as

$$| s_{11} | = \gamma | T_3 (tx) |, \quad x = \cos \theta \tag{5.11}$$

where γ is a real coefficient defining the mismatch tolerance, t is a scale factor chosen according to the specified bandwidth, and $\theta = 2\pi l/\lambda_g$ is the electrical length of a step.

The equivalent circuit of the stepped-impedance transformer in question along with relevant notation and the general shape of the desired reflection characteristic are shown in Fig. 5.16. The charac-

teristic impedances of the lines are designated as z_{c1} and z_{c5}, and those of the steps as z_{c2}, z_{c3} and z_{c4}.

At every joint of two transmission lines, there appears a reflected wave travelling to the left, for which the partial reflection coefficient is

$$\rho_i = (z_{ci} - z_{ci+1})/(z_{ci} + z_{ci+1})$$

Suppose that the change in the characteristic impedance at each joint is small. Then we may approximately write

$$\rho_i = \frac{z_{ci} - z_{ci+1}}{z_{ci} + z_{ci+1}} = \frac{\Delta z_c}{2z_{c\,av}} \approx \frac{1}{2}\Delta\{\ln z_c\} = \frac{1}{2}\ln\frac{z_{ci}}{z_{ci+1}} \qquad (5.12)$$

Because the reflection coefficients of the individual steps are small, we may well use what is known as the *first-approximation theory* under which it is assumed that the wave reflected from each joint does not undergo a change on passing through the remaining joints. The overall reflection coefficient of a stepped-impedance transformer under this assumption is equal to the sum of the reflection coefficients of the individual joints transferred to the start of the transformer:

$$\dot{s}_{11} = \rho_4 + \rho_3 \exp(-j2\theta) + \rho_2 \exp(-j4\theta) + \rho_1 \exp(-j6\theta)$$

It can be shown that the Chebyshev response defined by Eq. (5.11) holds when the reflection coefficients are symmetrical about the middle of the transformer. Therefore, $\rho_4 = \rho_1$ and $\rho_3 = \rho_2$, so that

$$\dot{s}_{11} = 2[\rho_1 \cos 3\theta + \rho_2 \cos\theta] \exp(-j3\theta)$$

On setting $x = \cos\theta$ and recalling the well-known forms for Chebyshev polynomials

$$T_1(x) = x = \cos(\text{arc}\cos x) = \cos\theta$$

and

$$T_3(x) = 4x^3 - 3x = \cos(3\,\text{arc}\cos x) = \cos 3\theta$$

we may write the magnitude of the reflection coefficient at the input of the transformer as

$$|\dot{s}_{11}(x)| = 2|\rho_1 T_3(x) + \rho_2 T_1(x)| = 2|4\rho_1 x^3 - 3\rho_1 x + \rho_2 x| \qquad (5.13)$$

On equating Eqs. (5.11) and (5.13), we get

$$\gamma(4x^3t^3 - 3xt) = 2[4\rho_1 x^3 - (3\rho_1 - \rho_2)x]$$

and find then the required partial reflection coefficients:

$$\rho_1 = \gamma t^3/2$$

and

$$\rho_2 = 3\gamma(t^3 - t)/2 \qquad (5.14)$$

In designing a stepped-impedance transformer, the coefficient γ and the scale factor t cannot be specified independently. The point is this. If we take the sum of the magnitudes of the reflection coefficients for the individual joints subject to Eq. (5.12), then, in view of (5.14), we shall obtain for $x = 1$

$$\sum_{i=1}^{4} \rho_i = \frac{1}{2} \ln (z_{c1}/z_{c5}) = 2 (\rho_1 + \rho_2) = \gamma T_3 (t)$$

Hence it follows that γ and t are interrelated

$$\gamma = \left| \frac{\ln (z_{c1}/z_{c5})}{2T_3 (t)} \right| \tag{5.15}$$

The scale factor t is solely determined by the specified bandwidth of match, $2\Delta\omega_m = \omega_2 - \omega_1$. To demonstrate, as follows from Eq. (5.11), the following conditions must be satisfied at the edges of the match bandwidth:

$$\left.\begin{aligned} t \cos \theta_2 &= -1 \\ \text{and} \qquad t \cos \theta_1 &= 1 \end{aligned}\right\} \tag{5.16}$$

where

$$\theta_1 = \omega_1 l/v_p$$

and

$$\theta_2 = \omega_2 l/v_p$$

From a closer look at the expression

$$\cos \theta_1 + \cos \theta_2 = 2 \cos \frac{\theta_1 + \theta_2}{2} \cos \frac{\theta_2 - \theta_1}{2} = 0$$

derived by use of Eq. (5.16), it is an easy matter to notice that at the centre (midband) frequency, $\omega_0 = (\omega_1 + \omega_2)/2$, the factor $\cos [(\theta_1 + \theta_2)/2]$ must vanish. Hence the length of each step must be a quarter of the wavelength in guide. From (5.16) it also follows that

$$\cos \theta_1 - \cos \theta_2 = 2 \sin \frac{\theta_2 - \theta_1}{2} \sin \frac{\theta_1 + \theta_2}{2} = 2/t$$

from which we can readily find the scale factor from the specified bandwidth of match:

$$t = \{\sin [(\pi/4) (2\Delta\omega_m/\omega_0)]\}^{-1} \tag{5.17}$$

The design procedure for the transformer must be as follows. The ratio of characteristic impedances z_{c1}/z_{c5}, the bandwidth of match $2\Delta\omega_m$, and the centre (midband) frequency ω_0 are specified in advance. The first step is to find the scale factor t and the mismatch γ over the bandwidth, using Eqs. (5.17) and (5.15). The next step is to determine the reflection coefficients ρ_1 and ρ_2 by Eq. (5.14). The final step is to determine the required characteristic impedances for the steps. If the mismatch γ over the specified bandwidth turns

out to be excessive, the number of impedance steps in the transformer should be increased.

A Butterworth (maximally flat) stepped-impedance transformer can be designed in much the same way as a Chebyshev (equal-ripple) stepped-impedance transformer, except that the reflection-vs-frequency response is specified in the form

$$| s_{11}(x) | = \gamma | t^n x^n |, \quad x = \cos\theta \tag{5.18}$$

With $n = 3$, this leads to the following relations:

$$\rho_1 = \gamma t^3/8$$

$$\rho_2 = 3\gamma t^3/8$$

$$\gamma = \left| \frac{\ln (z_{c1}/z_{c5})}{2t^3} \right|$$

Stepped-impedance transformers for which the response is defined by Eq. (5.18) are often called *binomial transformers* because the distribution of the reflection coefficients at the joints within the

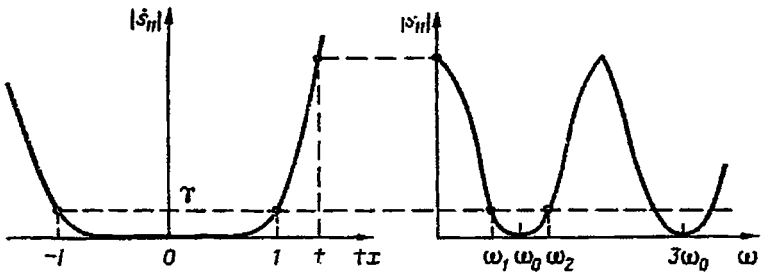


Fig. 5.17. Attenuation characteristics of a binomial transformer

transformer is described by a binomial expansion of degree n. The typical frequency responses of a binomial transformer are shown in Fig. 5.17.

In working out practical designs of stepped-impedance transformers, there is usually no need to perform all the steps of the procedure outlined above — the necessary data can be looked up in reference tables.

Apart from their use as broadband matching elements, stepped-impedance transformers can serve as prototypes from which to derive microwave filters, and also as multistep coupled-line directional couplers. For example, a stepped-impedance transformer serves as the prototype in a very successful design technique for direct-coupled microwave filters. To obtain a good frequency selectivity, the fractional change in characteristic impedance that a prototype transformer should produce may have to be several tens or even hundreds of thousands. Because of this the first-approximation theo-

ry which assumes that the individual reflections are small proves inapplicable. Instead, one has to use a more rigorous, but a more tedious synthesis theory. The equivalence between the prototype transformer and the direct-coupled filter being synthesized is obtained by identifying the transmission-matrix elements of the individual sections of the synthesized filter with the corresponding sections of the prototype.

5.8. Broadband Matching of Resistive Loads by Tapered-Impedance Transformers

A tapered-impedance transformer (Fig. 5.18) may be treated as the limiting case of a stepped-impedance matching transformer in which the number of steps is increased without bound and the length

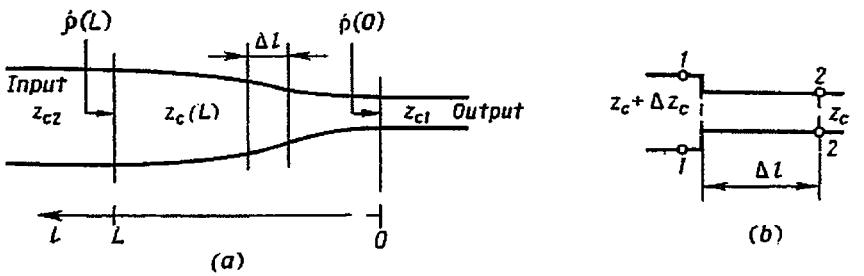


Fig. 5.18. Tapered matching transformer: *a*, general arrangement; *b*, elementary section

of each step tends to zero. At any transverse plane of a tapered-impedance transformer the reflection coefficient $\dot{\rho}(l)$ can be described by a differential equation of the form

$$d\dot{\rho}(l)/dl + j2\beta\dot{\rho}(l) + [1 - \dot{\rho}^2(l)]N(l) = 0 \qquad (5.19)$$

where

$$\beta = 2\pi/\lambda_g$$

is the phase constant assumed for simplicity to be independent of l, and

$$N(l) = \frac{1}{2}\frac{d[\ln z_c(l)]}{dl} = z_c'(l)/2z_c(l)$$

is the *local-reflection function* accounting for variations in the characteristic impedance $z_c(l)$ with distance along the transformer.

To derive Eq. (5.19) we should select inside the transformer an elementary step of length Δl, which produces a fractional change in characteristic impedance equal to $z_c/(z_c + \Delta z_c)$ (Fig. 5.18*b*). The

scattering coefficients of this step are

$$\left.\begin{aligned} \dot{s}_{22} &\approx -\dot{s}_{11} \approx \Delta z_c/2z_c = \tfrac{1}{2}\Delta\,[\ln z_c\,(l)] \\ \dot{s}_{12} &= \dot{s}_{21} \approx \exp(-j\beta\Delta l) \approx 1 - j\beta\Delta l \end{aligned}\right\} \tag{5.20}$$

The reflection coefficient at the left-hand port of the elementary step is

$$\dot{\rho}\,(l+\Delta l) = \dot{\rho}\,(l) + (\mathrm{d}\dot{\rho}/\mathrm{d}l)\,\Delta l = \dot{s}_{11} + \frac{\dot{\rho}\,(l)\,\dot{s}_{12}\dot{s}_{21}}{1-\dot{s}_{22}\dot{\rho}\,(l)}$$

In the last equality, Eq. (4.17) is used. On substituting (5.20) into this expression and neglecting the second-order infinitesimals as Δl tends to zero, we obtain the differential equation (5.19). An exact solution of this equation is difficult to obtain because it contains the squared function of the reflection coefficient. However, proceeding from the assumption that the transformer is a tapered one, we may take it that $|\dot{\rho}^2(l)| \ll 1$ and drop the corresponding term in the differential equation. The simplified equation

$$\mathrm{d}\dot{\rho}/\mathrm{d}l + j2\beta\dot{\rho} + N\,(l) = 0$$

is a first-order linear differential equation. Its solution can be obtained by invoking an integral factor and it has the form

$$\dot{\rho}\,(l) = \exp(-j2\beta l)\left[\dot{\rho}_0 - \int_0^l N\,(t)\exp(j2\beta t)\,\mathrm{d}t\right] \tag{5.21}$$

A similar equation can be derived within the framework of the first-approximation theory as the limit of the input reflection coefficient for a stepped-impedance transformer with its number of steps tending to infinity and with the length of each step tending to zero while the length of the entire transformer, L, is maintained constant.

The simplest example of a tapered-impedance transformer is the so-called *exponential transformer*. This is a nonuniform transmission-line section whose characteristic impedance varies as

$$z_c\,(l) = (z_{c1}z_{c2})^{1/2}\exp\left[\frac{2l-L}{2L}\ln(z_{c1}/z_{c2})\right]$$

On evaluating the local-reflection function

$$N\,(l) = (L/2)\ln(z_{c1}/z_{c2})$$

and substituting the result into Eq. (5.21), we obtain for $\dot{\rho}\,(0) = 0$

$$|\dot{\rho}\,(L)| = \tfrac{1}{2}\,|\ln(z_{c1}/z_{c2})|\;|\sin\beta L/\beta L|$$

With $\beta L \geqslant \pi$, that is, with $L/\lambda_g \geqslant 0.5$, the magnitude of the reflection coefficient does not exceed $0.1 \mid \ln (z_{c1}/z_{c2}) \mid$. As L/λ_g is increased, the magnitude of the input reflection coefficient of an exponential transformer decreases in the same manner as the level of side-lobe radiation in a linear equal-amplitude antenna on moving away from its main lobe (see Sec. 11.2). Thus, as regards the shape of the frequency response a tapered-impedance transformer is equivalent to a high-pass filter: a good match is obtained at all frequencies above some edge-point frequency.

The exponential transformer is an example of how Eq. (5.21) can be used to *analyse* the reflection coefficient when one knows the manner in which the characteristic impedance varies with distance along the transformer. It is far more difficult to *synthesize* optimal tapered-impedance transformers from the specified reflection-vs-frequency response. Here Eq. (5.21) becomes an integral equation in the unknown local reflection coefficient $N(l)$. Quite a number of worked-out examples for the synthesis of tapered-impedance transformers, including the Butterworth (maximally flat) and the Chebyshev (equal-ripple) types, can be found in reference books.

A comparison of stepped- and tapered-impedance transformers would show that, given the same fractional change in characteristic impedance and the same mismatch tolerance over the specified bandwidth, a stepped-impedance transformer will always be shorter than a tapered-impedance one. As already noted, a tapered-impedance transformer is equivalent to a high-pass filter and so the desired quality of match can be secured in a more economical way. Unfortunately stepped-impedance transformers are less robust electrically because the electric field concentrates at the step joints. When choosing a particular type of transformer, it is also important to consider other factors, such as the sensitivity of the match-vs-frequency response towards manufacturing tolerances, the quantity of materials needed, the cost, etc.

5.9. Broadband Matching of Complex Loads

The narrowband matching devices examined in Sec. 1.7 produce a pure travelling wave condition at a single, specified frequency. To achieve this condition, it suffices to use reactive matching elements with two degrees of freedom. Any deviation from the design frequency usually leads to an abrupt reduction in the TWR, and there is practically no possibility to control the frequency behaviour of narrowband matching elements.

In contrast to narrowband matching, the element values of broadband matching devices can be selected so that any one of two goals can be achieved:

(1) a maximum value for the TWR within the specified bandwidth;

(2) a maximum bandwidth of match for the specified value of the TWR.

The quality of broadband matching depends on how complicated the matching device is allowed to be. Also, for better results a reactive matching element must have more than two degrees of freedom. An algorithm for broadband matching must be developed with

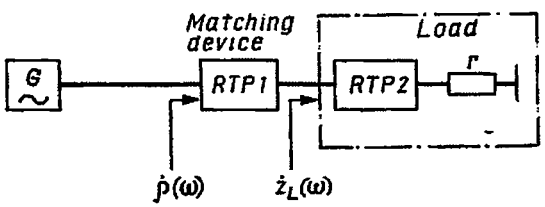


Fig. 5.19. General arrangement of a broadband matching network for a complex load

allowance for likely variations in load impedance over the bandwidth of interest, whereas for narrowband matching it is sufficient to know the value of load impedance at the centre (midband) frequency.

For the first time the problem of broadband matching involving lumped-element complex loads was formulated and solved by P.M. Fano of the United States in 1950. He showed that even if a reactive matching device has an infinite number of degrees of freedom, a pure travelling wave condition cannot be obtained over a finite continuous band of frequencies and that not just any complex load can be matched over the specified frequency band at the specified limit of TWR.

Fano's findings are based on an analysis of power transfer from a matched generator via an adjustable reactive two-port (labelled *RTP-1* in Fig. 5.19) to a complex load representable by a cascade connection of an arbitrary but fixed reactive two-port (labelled *RTP-2* in the same figure) and a fixed resistance r.

The fact that the equivalent circuit representation of the termination includes a fixed reactive two-port is responsible for constraints on broadband matching. If this two-port has a circuit which stops some frequency ω_s, no matching device will be able to provide for power transfer to the termination at that frequency. Examples of simple band-stop circuits are shown in Fig. 5.20. As is seen, any reactive element (that is, one capable of storing electromagnetic energy) will act as a stop circuit for a certain frequency. As Fano showed, the possibility of a good match away from the 'stop' frequencies is subject to constraints which take the form of a set of

integral inequalities

$$\int_{0-}^{\infty} \ln[1/|\dot{\rho}(\omega)|]\, F_i(\omega)\, d\omega \leqslant A_i, \quad i = 1, 2, 3, \ldots, N \qquad (5.22)$$

where $\dot{\rho}(\omega)$ = input reflection coefficient of the matching element
$F_i(\omega)$ = weighting functions (known functions of frequency) whose form depends on the structure of two-port *RTP-2* in the equivalent circuit of the termination
A_i = constant coefficients determined by the element values of *RTP-2* and the resistance of resistor r
N = number of integral inequalities; it is directly related to the number of independent reactive elements in the equivalent representation of the termination

In a geometrical interpretation, the inequalities in (5.22) are constraints on the areas under the function $\ln [1/|\dot{\rho}(\omega)|]$ subject

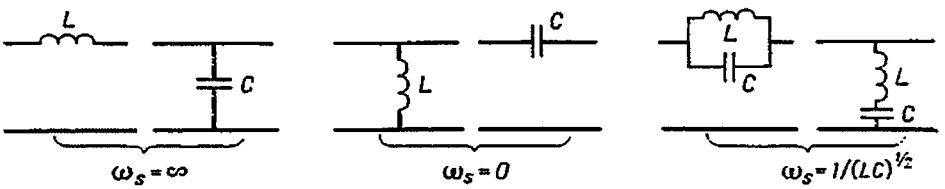


Fig. 5.20. Elementary stop-band two-ports

to variations in the weighting functions $F_i(\omega)$. A change in the response of the reactive matching device may substantially modify the form of the function $\dot{\rho}(\omega)$. However, the areas it bounds, when taken with a weight $F_i(\omega)$, cannot exceed the values of A_i.

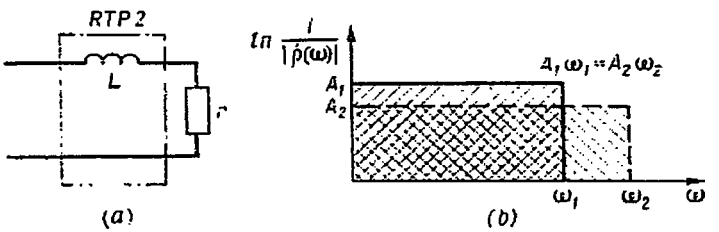


Fig. 5.21. Elementary complex load: *a*, low-pass prototype; *b*, optimal matching characteristics

For a simple complex termination representable as a series rL-circuit (Fig. 5.21*a*), the constraints in (5.22) reduce to an only

inequality

$$\int_0^\infty \ln\left[\frac{1}{|\dot{\rho}(\omega)|}\right] d\omega \leqslant \pi r/L \qquad (5.23)$$

Proceeding from (5.23), it is possible to establish the form of the optimal frequency response, $|\dot{\rho}(\omega)|$, which will give a maximum bandwidth of match with a TWR equal to the allowable value of TWR,

$$K_a = (1 - |\dot{\rho}_a|)/(1 + |\dot{\rho}_a|)$$

Obviously, within the bandwidth of match extending from zero frequency to the band-edge frequency ω_e, the function $\ln [1/|\dot{\rho}(\omega)|]$ must remain unchanged in value and equal to $\ln [1/|\dot{\rho}_a|]$, but outside that bandwidth it should be equal to zero. Such optimal characteristics are shown in Fig. 5.21*b* for two values: $\rho_{a1} < \rho_{a2}$. Assuming a rectangular shape for the optimal frequency response, the integral in (5.23) can readily be evaluated, and the maximum attainable bandwidth of match is found to be given by

$$\omega_e = \frac{\pi r}{L \ln [(1+K_a)/(1-K_a)]}$$

where K_a is the minimum allowable travelling-wave ratio.

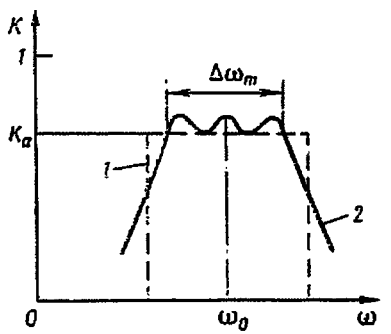


Fig. 5.22. Optimal TWR-vs-frequency characteristics under conditions of match: *1*, ideal; *2*, real under the Chebyshev approximation

By the change of the frequency variable, as defined in Eq. (5.5), this result can readily be extended to the matching of a resonant termination in the form of a series resonant circuit having a certain unloaded Q-factor, $Q_u = \omega_0 L/r$, where ω_0 is the resonance frequency. Assuming that the frequency dependence of the TWR in the vicinity of the resonance frequency has an ideal rectangular shape, we find for the bandwidth of match, $\Delta\omega_m$ (Fig. 5.22), that the relative bandwidth of match is

$$\Delta\omega_m/\omega_0 \leqslant \frac{\pi}{Q_u \ln [(1+K_a)/(1-K_a)]}$$

The above relation shows that the maximum attainable relative bandwidth of match, $\Delta\omega_m/\omega_0$, decreases with increasing unloaded Q-factor of the termination, Q_u, and with increasing allowable TWR. For example with $Q_u = 10$ and $K_a = 0.7$, the upper band-edge frequency is about 18.5%. With $K_a = 0.85$, it decreases to 11%. Of course, the

above equation holds also for a termination in the form of a parallel resonant circuit.

Analysis of the constraints imposed on the attainable bandwidth of match in the case of complex termination shows that, *in working out broadband matching devices, the designer ought not to strive for a perfect match at one or several points within the specified bandwidth.* This is because the presence of frequencies where $\rho = 0$ implies an uneconomical utilization of the limited values of the integrals in the relations (5.22). As the number of points of a perfect match within the specified bandwidth increases, the "dips" in TWR between them may be expected to become progressively deeper. Indeed, for proper broadband matching, it will be well-advised to spread the mismatch across the required bandwidth as evenly as practicable and to make the mismatch as large as possible outside the band.

The synthesis of broadband matching circuits is to a large extent similar to that of filters and is usually carried out with resort to the Chebyshev (equal-ripple) attenuation characteristic. One distinction consists in that some of the reactive elements in the matching circuit being synthesized belong to the termination and will therefore have fixed nominal values. The values of the other reactive elements in the matching circuit may be chosen such that they would complete the equivalent representation of the termination to a Chebyshev (equal-ripple) filter (quite possibly, with a further mismatch, as shown in Fig. 5.22).

Chapter Six

Microwave Control and Ferrite Devices

6.1. Classification of Microwave Control Devices

The term 'control' in the present context refers to microwave devices which serve to control the amplitudes and phases of microwave signals or to change the direction of propagation in microwave transmission systems.

Control devices can be *mechanical* or *electrical.* In mechanical devices, the desired change is achieved as a result of displacements of individual parts. In electrical devices, the desired change is brought about in response to electric signals and does not entail any mechanical motion.

Electrical control devices for microwave applications can be based on microwave switching diodes, biased ferrites, gas-discharge tubes, and ferroelectric elements.

A further classification of microwave control devices is according to the functions they are intended to perform.

1. *Switches.* These are four-terminal devices which either provide for a matched transmission of a signal from input to output (the ON state) or block signal transmission (the OFF state). The OFF state can be effected by either completely reflecting the signal or completely absorbing it.

2. *Commutators.* These are multi-terminal devices having one or more inputs and a number of outputs. Signals applied to the inputs are conveyed over one or several switchable paths to the outputs with minimum losses and in compliance with match requirements.

3. *Reflecting phase shifters.* These are one-ports whose reflection coefficient has a magnitude close to unity and an adjustable phase. Ideally, a reflecting phase shifter is a variable reactance and can be implemented as a short- or open-circuited stub of a variable length.

4. *Feed-through phase shifters.* These are matched two-ports. Ideally, they are reactive in behaviour and have a variable phase of the transmission coefficient. They can be continuous (with the phase variable continuously) and stepped (with a phase that can take on a series of discrete values differing by a constant increment, $\Delta\varphi$).

5. *Attenuators.* These are two-ports which introduce a loss which can be varied continuously or in steps from a minimum value, L_{min}, to a maximum value, L_{max}.

6. *Power limiters.* These are self-adjusting match-terminated two-ports which transmit weak signals almost without attenuation but limit signals whose power exceeds some predetermined level.

6.2. Mechanical Switching Devices. Phase Shifters. Attenuators

Mechanical switching devices are a regular feature of waveguide systems. A simple switch (Fig. 6.1*a*) couples two guides together by means of a waveguide channel made in a metal rotor. For better electric contact, the rotor has on its outer surfaces several choke grooves similar to those made in an ordinary choke flange coupling. When the rotor is turned through 90°, the interconnected guides are shorted, and the choke grooves prevent the signal from leaking around the gap between the rotor and the housing. In a more elaborate switch (Fig. 6.1*b*), the rotor has three waveguide channels so that four ports can be connected in four distinct patterns.

In *mechanical non-reflecting phase shifters* the electrical length of the transmission-line sections can be varied either by adjusting their physical length or by varying the phase coefficient. As an example, we shall dwell on a mechanical phase shifter known as the *trombone.* This type of phase shifter intended for use in rigid coaxial guides is shown in Fig. 6.2. For better matching the sliding contacts on the inner and outer conductors are spaced apart along the guide in such a way that the diameter ratio of the inner to outer conductor is the

same at all sections being joined, thereby giving a constant characteristic impedance. Reflections likely to occur at the joints are further suppressed by offsetting the steps, Δl, on the inner and outer conductors.

Figure 6.3 shows *waveguide phase shifters* which involve variations in the phase coefficient of the TE_{10} mode in rectangular guide. In

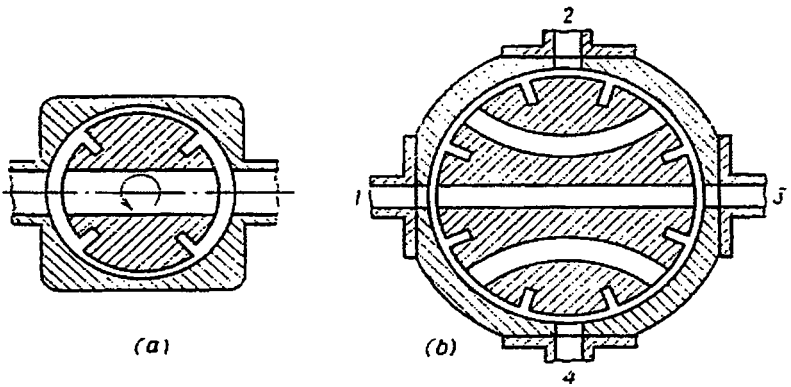


Fig. 6.1. Mechanical waveguide switches

the design of Fig. 6.3*a*, a dielectric strip is extended into the rectangular guide through a non-radiating longitudinal slot in the middle of the broad dimension parallel to the electric lines of force. As a result, the wave is slowed down and the delay at the output of the phase shifter is increased. The tapered ends of the dielectric strip improve the quality of match for the phase shifter. By varying the taper, it is possible to control the manner in which the resultant phase shift depends on the depth of immersion of the dielectric strip. In the design of Fig. 6.3*b*, the dielectric strip is moved from the edge towards the centre of the guide by a holder which extends through a hole in the narrow dimension (side face) of the guide. As the strip is moved deeper into a region of high concentration of the electric field, the wave in the guide is slowed down increasingly more, and the delay introduced by the phase shifter is likewise increased. The ends of the strip are tapered to reduce reflections.

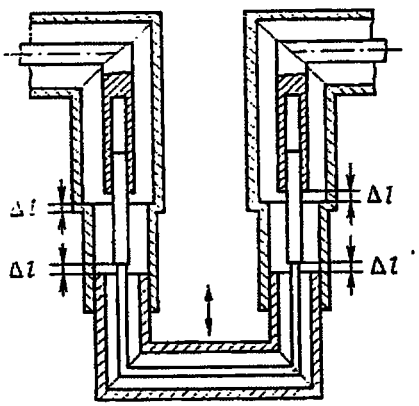


Fig. 6.2. Coaxial phase shifter of the trombone type

In the design of Fig. 6.3*c*, the waveguide is made of a flexible material, and non-radiating slots are cut in the middle of the broad faces. When an external force F is applied to the side faces, the guide width a is changed, and this leads to a change in the cutoff wavelength, $\lambda_c = 2a$, and in the phase velocity of the TE_{10} mode (the contraction of the guide reduces the phase delay).

Variable attenuators utilize the exponential attenuation of modes in below-cutoff waveguides. Figure 6.4 shows several designs of

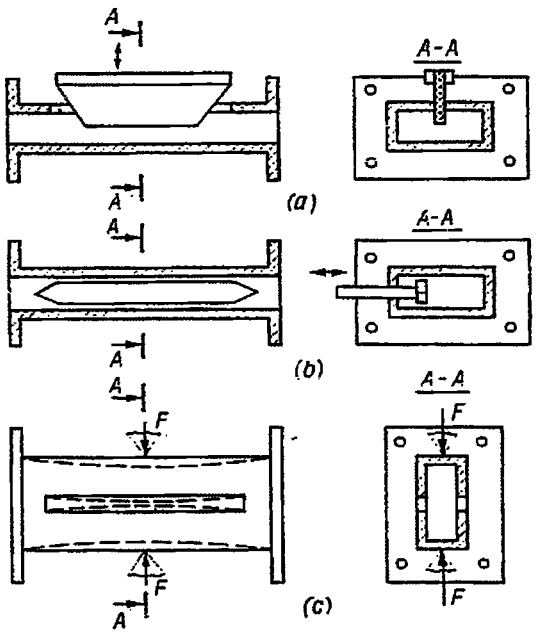


Fig. 6.3. Mechanical waveguide phase shifters

coaxial variable attenuators fabricated in circular guide of variable length and small diameter, in which all modes are non-propagating. In the design of Fig. 6.4*a*, a TM_{01} mode is launched in a circular below-cutoff guide by a capacitive probe, whereas in the design of Fig. 6.4*b*, a TE_{11} mode is launched by a coupling loop. At the other end, the attenuated wave is picked up by a capacitive probe or a coupling loop connected to a coaxial waveguide.

The attenuation (in dB) introduced by a *below-cutoff waveguide attenuator* is directly proportional to the length of the below-cutoff waveguide section, so it can be varied between broad limits.

The attenuators shown in Fig. 6.4*a* and *b* are not matched, so the attenuation they provide is due to reflections from the ports.

The desired match in below-cuttoff attenuators is obtained by introducing absorbers. Fig. 6.4*c* shows a matched below-cutoff waveguide attenuator in which the absorbers are resistors. The central conductor of the incoming coaxial line terminated to a matched load R_1 launches in the circular guide of the side arm a decaying TE_{11} mode which is picked up at the other end of the guide by an inductive probe. The probe circuit includes a resistive absorber R_2 which

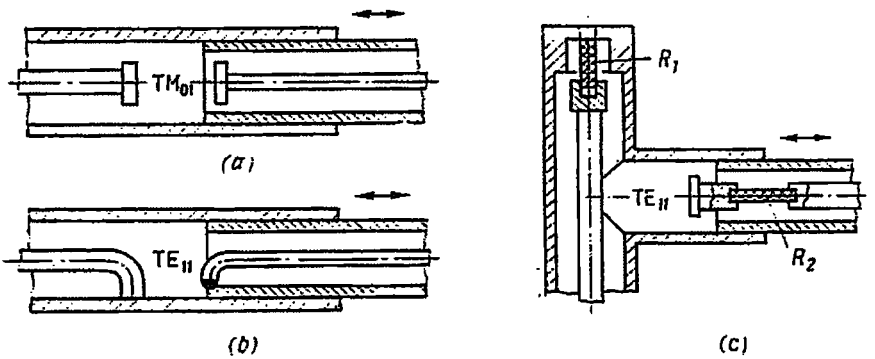


Fig. 6.4. Below-cutoff coaxial waveguide attenuators

ensures a proper match at the output of the attenuator. The amount of attenuation is varied by changing the length of the circular guide with a sliding contact in the outer conductor of the outgoing coaxial guide.

A further form of attenuators used in waveguide systems is a waveguide attenuator enclosing an absorbing strip. In design, the *strip attenuator* is not unlike the waveguide phase shifters shown in Fig. 6.3, but the dielectric strip is replaced by an absorbing material. As the absorbing strip is moved into a region of high electric field intensity, more power is dissipated in the strip, and the amount of attenuation due to the attenuator is increased. The shape of the strip controls both the quality of match and the calibration of the attenuator.

6.3. TR and ATR Tube Duplexers

In pulsed radar equipments that use the same antenna for both transmission and reception some form of switch is needed. During transmission it should connect the antenna to the transmitter and, at the same time, disconnect the receiver from the antenna in order that the transmitted pulse could not do permanent damage to the receiver circuits. During reception it should disconnect the transmitter from the antenna because otherwise the transmitter would absorb an appreciable proportion of the target echo signal and reduce the

power delivered to the receiver, thereby reducing the maximum detection range of the radar.

This double-acting switch is quite aptly called a *duplexing assembly* or, simply, a *duplexer*. It commonly includes gas-filled switching tubes mounted within waveguide segments, called a transmit-receive (TR) tube, and an anti-transmit-receive (or ATR) tube.

Figure 6.5 shows a resonant waveguide TR tube used in most duplexing assemblies. It is a reflecting switch in the form of a her-

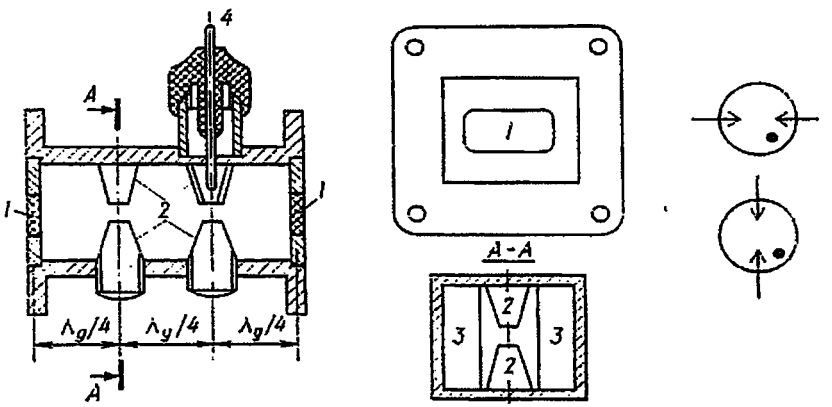


Fig. 6.5. Resonant TR tube and its symbols: *1*, resonant windows; *2*, truncated-cone electrodes; *3*, inductive diaphragm; *4*, ignitor (keep-alive) electrode

metically sealed rectangular waveguide filled with a mixture of water vapour and argon or hydrogen at a low pressure (of the order of 10^2 to 10^4 Pa). At both the input and output it has resonant diaphragms, or windows, sealed with glass or ceramic plates transparent to microwave energy. Arranged along the centre line of a broad guide face, within a quarter of a guide wavelength of each other and of the resonant windows, are truncated-cone electrodes which act as capacitive posts in the absence of a gas discharge. The segments taken up by the truncated cone electrodes also contain matching inductive diaphragms that complete each electrode structure to one having the equivalent circuit of a parallel resonant circuit. For weak passing signals the TR tube acts as a quarter-wave-coupled band-pass filter with a relative bandwidth of 5 to 10%. The tube is switched to reflect the signal automatically under the action of the strong transmitter pulse. So as to keep to a minimum the level at which this can happen, an *ignitor* or *keep-alive electrode* structure is placed within one truncated cone — it acts to maintain a d.c. glow discharge in the gap. Arrival of a strong microwave signal produces a discharge between the cones enclosing the ignitor electrode, then the gap breaks down between the other pair of truncated-cone electrodes located in

a loop of the field produced by a standing wave. Finally, a discharge occurs which produces an effective short-circuit for the microwave signal in the plane of the input window. Most TR tubes break down and become fully conducting in a matter of 10^{-8} s after receiving the excitation, and they recover after the microwave pulse ceases in roughly 10^{-5} to 10^{-6} s, when the previously ionized gas in the gap has been deionized.

At ultra-high and lower frequencies (decimetric and longer wavelengths), demountable rather than integral TR and ATR tubes are

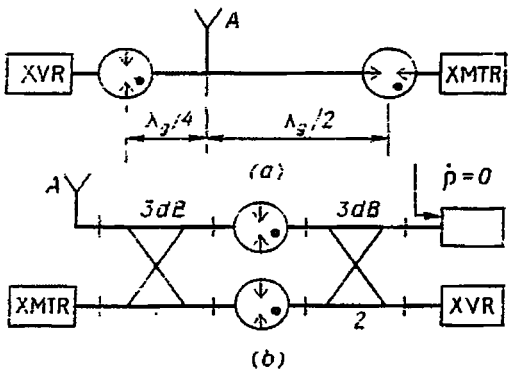


Fig. 6.6. Duplexers: *a*, tapped type; *b*, balanced type

used in radar duplexers. They, too, are *gas-filled, glass-envelope spark gaps* that are installed in series or in parallel with the transmission line between the antenna and the transmitter and receiver. Cooperating with suitable passive elements within the microwave transmission system, they form a device equivalent to the integral resonant TR-ATR tube just discussed.

In general, TR and ATR tubes can be combined in two different duplexer configurations, namely *tapped* and *balanced*. A tapped duplexer (Fig. 6.6*a*) contains a TR tube in the side arm leading to the receiver and a series-connected ATR tube in the transmitter arm. To avoid interaction between the transmit and receive channels, the TR and ATR tubes are placed within a quarter and a half of the wavelength in the guide from the side-arm, respectively. Upon arrival of a transmitter pulse, the two gaps break down and present an effective short-circuit. The transmitter pulse is now free to go to the antenna, whereas the receiver is isolated by the TR tube and its channel acts as a quarter-wave isolator. In between transmitter pulses the ATR tube is open-circuited and its high impedance is transformed by a half-wave line segment into an open-circuit at the parallel junction point, so that the transmit channel cannot shunt the receive channel.

Among the demerits of tapped duplexers are the need to use different gas-discharge tubes, a marked reduction in power during transmission, an inadequate isolation of the receiver from the transmitter, and a limited bandwidth.

These drawbacks are avoided in *balanced duplexers* (Fig. 6.6*b*). This type of duplexer uses two 3-dB directional couplers with their decoupled ports interconnected pairwise by transmission-line (guide) segments containing each a resonant spark-gap tube. If we trace the signal flow through a balanced duplexer in both the broken-down and

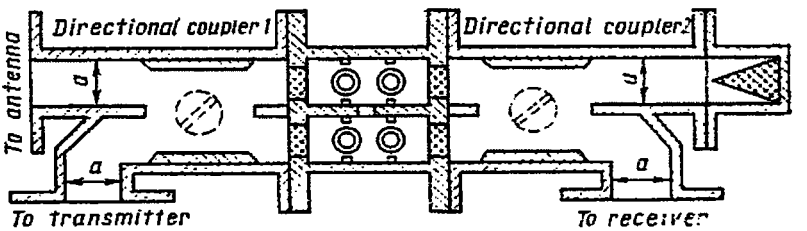


Fig. 6.7. Dual TR tube balanced duplexer

recovered state of the spark-gap tubes and recall the properties of 3-dB directional couplers (see Examples 4.6 and 4.7), we will see that:

(1) the transmitter pulse power is divided into two equal parts by the first directional coupler and causes the spark-gap tubes to fire (that is, their gaps to break down). The pulses reflected from the gaps travel through the first coupler again and are combined in the antenna. At the input of the first coupler connected to the transmitter the reflected pulses arrive in anti-phase and cancel out so that no reflected waves exist in the transmitter channel;

(2) the energy leaking through the resonant spark-gaps during transmission is combined in the second coupler match-terminated at the output, and cancels out at the port connected to the receiver;

(3) when the transmitter is not sending out, the echo signals picked up by the antenna are free to pass through the spark-gap tubes and are combined in the receiver. The signals reaching the ballast termination are in anti-phase — they cancel out, and no power is lost during reception;

(4) during reception the transmitter is isolated from the receiver owing to the action of the directional couplers. Because of the directivity of the couplers balanced duplexers improve isolation between the receiver and the transmitter by 7 to 10 dB.

In sketch form, a balanced duplexer using two 3-dB directional couplers is shown in Fig. 6.7. The two sections of the dual spark-gap tube are filled by the same gas owing to an opening in the common

wall. The two channels of the dual spark-gap tube have an identical performance — a fact which improves the operation of the duplexer during both transmission and reception.

6.4. Microwave Switching Diodes

Microwave switching diodes are designed for a peak power of up to 100 kW in pulsed work and an average power of up to 1 kW in CW work.

Most often, microwave switching diodes are of the *PIN* (or *p-i-n*) type (Fig. 6.8). The structure of a *PIN* diode consists of a heavily doped *p*+-region and a heavily doped *n*+-region separated by

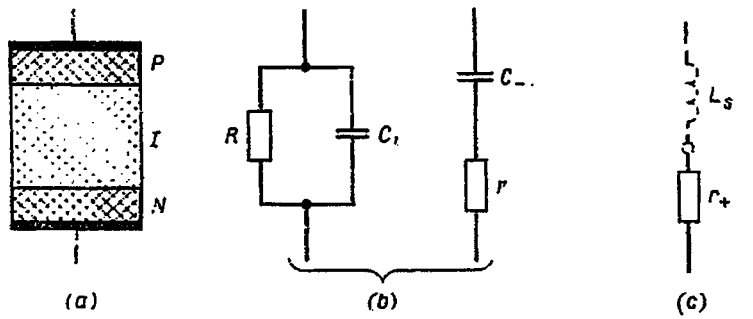


Fig. 6.8. PIN diode and its equivalent circuit: *a*, general arrangement; *b*, diode in OFF state; *c*, diode in ON state

a high-resistance, lightly doped intrinsic (or *i*-) layer usually called the *base* of a *PIN* diode. The areas on the diode ends, about 1 mm in diameter, adjacent to the *p*- and *n*-regions, are metallized and used as terminals. When a *PIN* diode is at zero or reverse bias, the contact potential differences across the *p-i* and *i-n* junctions prevent migration of free carriers from the *p*- and *n*-regions into the base, and the diode presents a very high resistance (units or even tens of kilohms). Because the *i*-region is usually very thick (several hundred micrometers), *PIN* diodes suffer from a marked time lag.

When microwave energy is applied to a reverse-biased *PIN* diode, no rectification takes place because free carriers have no time to accumulate in the base (*i*-) region during the positive half-cycle. At zero or reverse bias a *PIN* diode can stand up to microwave voltages as high as 10^3 volts without displaying any nonlinearity. Therefore, the equivalent circuit of a reverse-biased *PIN* diode (Fig. 6.8*b*) is either a parallel connection of a large resistance R (with a value of several kilohms) and the overall diode capacitance C_t (equal to 0.3 to 1 pF) or a series connection of a small loss resistance r_- (with

a value of several ohms) and the base capacitance $C_{\pm}$. When a *PIN* diode is forward biased by a positive voltage of 1 to 2 volts, the junctions are rendered conducting, and carriers are injected into the base — holes from the *p*-region and electrons from the *n*-region, and the base resistance drops abruptly. To maintain the base resistance at a low value, a direct current of 1 to 200 mA is required so as to make up for the loss of carriers due to recombination.

A forward-biased *PIN* diode is capable of passing microwave currents up to 100 A while retaining a low resistance because the negative half-cycle of microwave energy has no time to withdraw

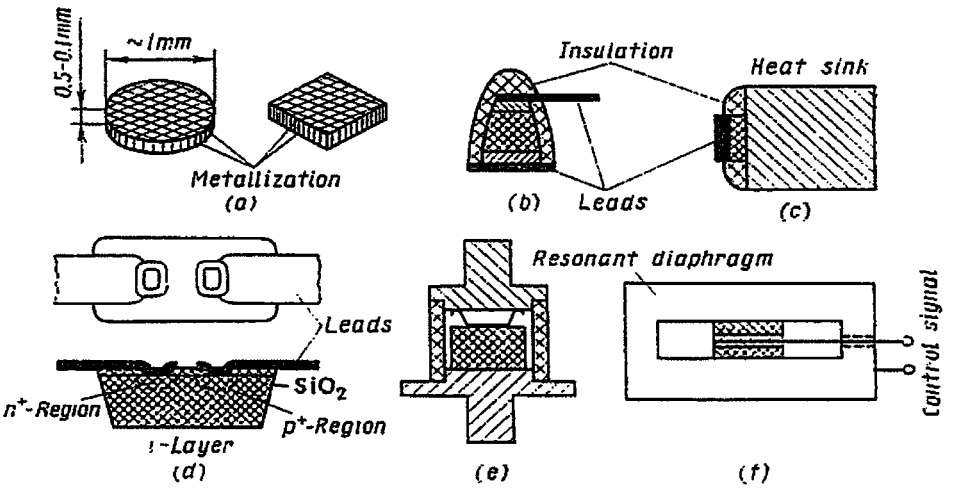


Fig. 6.9. Several designs of PIN diodes: *a*, *b*, unpackaged diodes; *c*, diode with a metal heat-sink; *d*, surface-oriented diode; *e*, PIN diode in a cermet package; *f*, twin PIN diode in a resonant diaphragm

any part of the space charge from the base. A forward-biased (conducting) *PIN* diode is representable by an equivalent circuit which consists of an only resistance, r_+ (with a value of several ohms). For a packaged *PIN* diode the equivalent circuit may sometimes be extended to include a low lead inductance, L_s (Fig. 6.8*c*).

PIN diodes are rather sluggish in their response to control signals. Their turn-on time, which is a function of the rate of injection of carriers into the *i*-region, is from 0.1 to 1 μs. Their turn-off time in which the principal component is the storage time (the time interval required for the excess minority carriers to be removed from the *i*-region) is much longer.

PIN diodes are fabricated from single silicon crystals by the diffusion or the ion-implantation processes. Several designs of *PIN* diode packages are shown in Fig. 6.9.

Apart from *PIN* diodes, microwave structures employ *P-N* switching diodes and varactors.

Impedance transformation by microwave switching diodes. Switching diodes (as well as any other control one-ports) are usually connected in microwave control circuits by means of impedance-transforming two-ports. The two possible states, *1* and *2*, of a diode correspond to the following complex normalized impedances:

$$\dot{z}' = r' + jx'$$

and

$$\dot{z}'' = r'' + jx''$$

which form, between them, an *impedance pair* $(\dot{z}', \dot{z}'')$.

Let an arbitrary switching one-port with an impedance pair $(\dot{z}', \dot{z}'')$ act as a load (termination) for some passive two-port which has an A matrix of the form

$$\mathrm{A} = \begin{bmatrix} \dot{a} & \dot{b} \\ \dot{c} & \dot{d} \end{bmatrix}$$

The two termination states, *1* and *2*, correspond to the input impedances of the two-port, given by Eq. (4.2)

$$\dot{z}_1' = (\dot{a}\dot{z}' + \dot{b})/(\dot{c}\dot{z}' + \dot{d})$$

and

$$\dot{z}_1'' = (\dot{a}\dot{z}'' + \dot{b})/(\dot{c}\dot{z}'' + \dot{d})$$

which form, between them, a *transformed impedance pair*, $(\dot{z}_1', \dot{z}_1'')$.

As has been found, the input and output impedance (or admittance) pairs of an impedance-transforming two-port display the following property: *The measure of difference for a pair of impedances (or admittances) defined as*

$$M_z(\dot{z}_i', \dot{z}_i'') = |\dot{z}_i' - \dot{z}_i''|/|\dot{z}_i' + \dot{z}_i''^*|$$

remains unchanged as the pair is transformed by an arbitrary reactive two-port.

To prove the above statement, it would suffice to use the A-matrices shown in Fig. 4.2 and verify that the measure M_z remains unchanged if the transformation is achieved by any elementary reactive two-ports, such as a series or a shunt reactance in a transmission line, a butt joint of two transmission lines, or an ideal quarter-wave transformer. Because at a given frequency a reactive two-port may be represented as a cascade connection of (at least) three elementary two-ports, the invariance of the measure M_z upon transformation by a reactive two-port is proved.

Now we set out to prove the following lemma: *Any switching one-port having an impedance pair* $(\dot{z}', \dot{z}'')$ *can be transformed at a specified frequency and by use of a reactive impedance-transforming two-port into a canonical switching element having a pair of pure resistances,* r *and* Kr, where $K > 1$ is the real *quality parameter* of the pair.

For purposes of proving the lemma, we will describe one of the methods for the synthesis of the transforming two-port. Connect a narrowband reactive matching device (see Sec. 1.7) to the switching one-port and choose two of its independent variables so as to ideally match one of the impedances of the pair, say, $\dot{z}'$. This will give rise to a pair of impedances $(1, \dot{z}_1'')$ at the input of the matching device, such that the impedance $\dot{z}_1''$ can be found by the transformation equation (4.2) in which the load impedance z'' and the parameters of the A matrix of the matching device should be substituted. Shift the phase reference plane at the input of the matching device to a voltage loop when the switching device is in the state z_1''. At the loop the pair of impedances will take the form $(1, K)$ where $K > 1$ is the standing wave ratio of the input line when the load is in the state $\dot{z}_1''$. Using a quarter-wave transformer with a characteristic impedance $r^{1/2}$, reduce the pair of impedances $(1, K)$ to the canonic form (r, Kr) formulated in the lemma. Thus, we have proved the lemma.

Since the reactive transforming two-port synthesized in proving the lemma preserves the value of the measure M_z, the following equality holds

$$(\dot{z}' - \dot{z}'')/(\dot{z}' + z''^*) = (K - 1)/(K + 1)$$

which, after identical transformations, reduces to the equation of the quality parameter

$$K = \frac{[(r' + r'')^2 + (x' - x'')^2]^{1/2} + [(r' - r'')^2 + (x' - x'')^2]^{1/2}}{[(r' + r'')^2 + (x' - x'')^2]^{1/2} - [(r' - r'')^2 + (x' - x'')^2]^{1/2}}$$

The above equation is convenient to use when evaluating the quality parameter from the measured impedances of the switching diode in the two operating states. Physically, the quality parameter is the standing wave ratio at the input of the reactive transforming two-port when the load is in state *2* on the condition that the two-port ideally matches the load in state *1*. A switching one-port with a purely reactive impedance at least in one of the two states has an infinite quality parameter.

For switching *PIN* diodes operating in the UHF and SHF bands, the quality parameter typically is 10^3 or 10^4. The quality parameter

is a universal index in terms of which any control elements can be compared. This parameter defines the lowest attainable level of attenuation in control devices built of nonideal switching elements.

6.5. Microwave Switches Using Switching Diodes

The simplest switch contains one switching element with a pair of resistances, r and Kr, connected in shunt or series with a transmission line (Fig. 6.10*a* and *b*). The value of r must be chosen such that $r \ll 1$ and $Kr \gg 1$.

The attenuation due to a shunt switch in the two states of the switching element is given by

$$L_{off} = 1/|\dot{s}'_{21}|^2 = (1 + 1/2r)^2$$
$$L_{on} = 1/|\dot{s}''_{21}|^2 = (1 + 1/2Kr)^2$$

where $\dot{s}'_{21}$ and $\dot{s}''_{21}$ are the scattering coefficients of the switch in the OFF and ON states of the switching element. In deriving the above

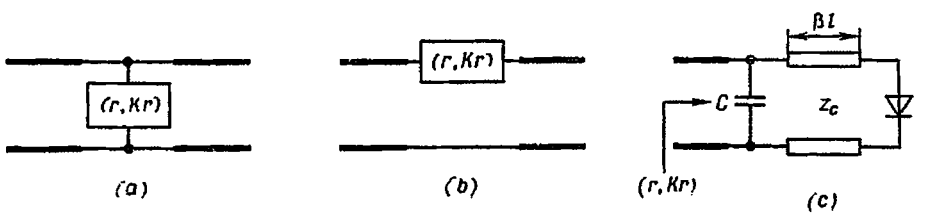


Fig. 6.10. Simple microwave diode switches: *a*, parallel type; *b*, series type; *c*, circuit of a switching element

equations use may be made of the last A matrix in Fig. 4.2 and of Eqs. (4.5) in order to convert the A matrix to the S matrix.

The attenuations L_{off} and L_{on} are interrelated

$$(L_{off}^{1/2} - 1)/(L_{on}^{1/2} - 1) = K \qquad (6.1)$$

Thus, the quality parameter of a switching element does set the limit of performance attainable by the associated switch. In design calculations, only L_{off} or L_{on} may be chosen at will; the other is uniquely related to the first by Eq. (6.1). The values of L_{off} and L_{on} can be adjusted by varying the value of r. For example, such a switch can set the limit for a maximum microwave power if the ohmic losses will be the same in each of the two states. It can be shown that this requirement will be satisfied with

$$r = 1/(2K^{1/2})$$

and the attenuations due to the switch optimized in such a manner will be

$$L_{off} = (1 + K^{1/2})^2$$

and

$$L_{on} = (1 + 1/K^{1/2})^2$$

With $K = 10^3$, this works out to $L_{off} = 30.3$ dB and $L_{on} = 0.27$ dB. Similar results are obtained for a series-connected switch as well.

A practical embodiment of such a switch for a TE_{10} rectangular guide is a resonant iris combined with a dual *PIN* diode (see Fig. 6.9*f*). In the ON state, the diodes short out the iris and drive the switch to the OFF state. In the OFF state of the diodes their capacitances are balanced out by a reduction in the iris opening, and the switch introduces a minimal attenuation. The fine conductor that controls the diodes runs along the opening and at right angles to the electric lines of force in the iris. Therefore, it has practically no effect on the performance of the switch at microwave frequencies.

The electric parameters of microwave switches (the attenuation introduced in each of the two states and the maximum power dissipation) can be improved through the use of more elaborate circuits containing several switching elements spaced apart along a transmission line.

Diode switches can successfully be used in duplexers (see Fig. 6.6) instead of or along with TR/ATR tubes.

6.6. Discrete Switching-Diode Phase Shifters

Reflecting phase shifters (RPS). The usual prototype for a two-state reflecting phase shifter is the circuit shown in Fig. 6.10*c* where the parameters of the reactive impedance-transforming two-port are chosen to safisfy the following requirements:

(1) The difference in phase between the reflection coefficients in the two states of the diode must be equal to the specified incremental change in phase:

$$\Delta\varphi = | \varphi_2 - \varphi_1 |$$

(2) The reflection coefficients ρ' and ρ'' in the two states of the diodes should be equal in magnitude and differ from unity as little as possible.

The two requirements define what may be called the *optimization of a phase shifter in terms of losses.*

Ohmic losses in a reflecting phase shifter are customarily stated in terms of the attenuation it introduces

$$L = 1/\rho^2$$

Analysis shows that when the reflection coefficients are the same in magnitude, $\rho' = \rho''$, the attenuation due to a reflecting phase shifter is

$$L = 1 + (4/K^{1/2}) \, | \sin (\Delta\varphi/2) | \qquad (6.2)$$

that is, it is independent of the parameters of the reactive impedance-transforming two-port and is solely decided by the specified incremental change in phase and the quality parameter of the diode. Note that canonical switching elements with a pair of resistances (r, Kr) may be treated as a reflecting phase shifter with the incremental change in phase equal to π.

Multi-state reflecting phase shifters are often constructed in the form of a transmission-line segment shunted by canonical switching elements at several sections (Fig. 6.11*a*). One of the switching

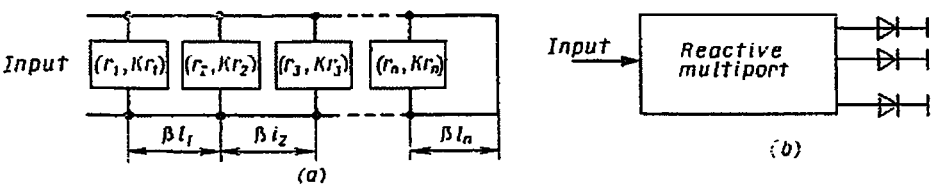


Fig. 6.11. Multi-state reflecting phase shifters

elements having a low resistance ($r \ll 1$) shorts the transmission line, while the remaining switching elements which have a high resistance ($Kr \gg 1$) have no effect on the phase of the reflection coefficient. When the switching elements change state, the location of the short-circuit plane on the line and the phase of the reflection coefficient are changed, too.

Other designs of multi-state reflecting phase shifters have been developed, all of them traceable to the prototype in Fig. 6.11*b*. In a reactive N-port junction, one of the ports is the input of the phase shifter, whereas the other $N - 1$ ports are terminated in switching diodes. The independent variables of the N-port junction are chosen so that each combination of conducting (ON) and non-conducting (OFF) diodes results in one of the discrete states of the phase of the reflection coefficient.

Feed-through diode phase shifters. These devices are required to give a specified difference in phase between the transmission coefficients

$$\Delta\varphi = |\varphi_2 - \varphi_1|$$

in the two states on the proviso that the ports are matched and the attenuation due to them is a minimum.

The simplest form of such a device is a *feed-through diode phase shifter using switchable transmission-line segments* (Fig. 6.12*a*). The phase of the transmission coefficient is incremented by

$$\Delta\varphi = \beta (l_2 - l_1)$$

as microwave energy is channelled by suitable diode switches to travel a distance l_1 or l_2. Analysis shows that the attenuation in such

a phase shifter using diodes for which the quality parameter is $K > 100$ depends only slightly on $\Delta\varphi$, is about the same in each phase state, and can be written as

$$L = 1/|\dot{s}_{21}|^2 \approx 1 + 4/K^{1/2}$$

That is, it corresponds to the maximum attenuation given by Eq. (6.2). Because of this, switchable line-segment phase shifters are unattractive where small phase shifts are needed.

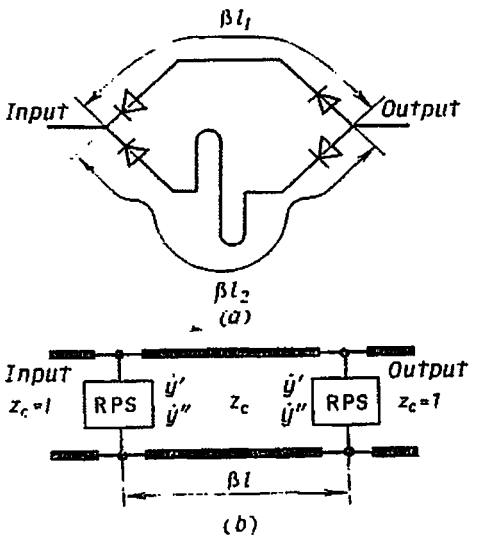


Fig. 6.12. Two-state non-reflecting phase shifters: *a*, switched line-section type; *b*, loaded-line type

Microwave transmission systems widely use *coupler-type feed-through phase shifters*. A coupler-type phase shifter is obtained by connecting two identical reflecting diode phase shifters to the mutually decoupled (isolated) output ports of a 3-dB directional coupler (see Fig. 4.20*a*). The operating principle of coupler-type phase shifters has been examined in Sec. 4.8 (see Example 4.7). The phase shift and attenuation due to a coupler-type phase shifter are functions of the phases and magnitudes of the reflection coefficients of the reflecting phase shifters used. When optimized reflecting phase shifters are used, the attenuation due to the composite coupler-type phase shifter is given by Eq. (6.2). Discrepancy in performance between the constituent *PIN* diodes and the non-ideal behaviour of the 3-dB directional coupler impair the quality of match in coupler-type phase shifters and lead to an excess attenuation and phase errors.

A loaded-line feed-through phase shifter (Fig. 6.12*b*). The key parameters of this compact device are the shunt admittances of the reflecting phase shifters, *RPS*, in the two states,

$$(\dot{y}', \dot{y}'') \approx j\,(b', b'')$$

the electrical length βl, and the characteristic impedance of the line segment between them. These parameters are chosen such that the input of the phase shifter will be matched in the two states and that the required incremental change in the phase of the transmission coefficient, $\Delta\varphi$, will be obtained. Of practical interest is a 'symmetrical' case, $b' = -b''$, when nearly the same passbands

are obtained in the two states and a good performance over the bandwidth is secured. The element values of the circuit are given by

$$b' = \tan(\Delta\varphi/2)$$

$$\beta l = \pi/2$$

$$z_0 = \cos(\Delta\varphi/2)$$

The attenuation due to the finite quality of the switching diodes

$$L = 1 + (4/K^{1/2}) \mid \tan(\Delta\varphi/2) \mid$$

shows a stronger dependence on the phase shift as compared with Eq. (6.2). (In the above equation, the sine is replaced by the tangent of the incremental change in phase.) Therefore, loaded-line feed-through phase shifters offer no advantage when used at large phase shifts ($\Delta\varphi > \pi/2$).

Multistate feed-through phase shifters. These devices can be built in more than one way. In coupler-type phase shifters the number of phase states can be increased through the use of multi-state reflecting phase shifters (see Fig. 6.11*a*). Switchable-line phase shifters may contain more than two switchable line segments.

Most often, a multistate feed-through phase shifter is built as a cascade connection of two-state feed-through phase shifters. If the cascaded stages are identical and their number is equal to the required number of phase states, the resultant circuit is non-optimal in terms of the overall attenuation it introduces, but it offers a major advantage which consists in that the power loss is split equally among the diodes, and the phase shifter transmits more power.

Fewer diodes need to be used in a multistate phase shifter built on the *binary principle*, that is, as a cascade connection of p feed-through stages the first of which shifts the phase by π, the second by $\pi/2$, the third by $\pi/4$, and so on. The total number of phase states is thus equal to 2^p, which gives a phase span from zero to 2π in steps decided by the stage giving the least phase increment. When properly optimized, a binary feed-through phase shifter will introduce the least possible attenuation, but its diodes will be operating under different conditions, and this reduces the maximum transmitted power. A binary phase shifter may use different types of individual two-state phase shifters. The stage with a phase shift of π or $\pi/2$ may be a directional-coupler type, and the stages with smaller phase shifts may be of the simpler loaded-line type.

The pattern produced on a microstrip board to obtain a three-state *PIN*-diode feed-through phase shifter is shown in Fig. 6.13. This is an example of a microwave hybrid integrated circuit. Here the term 'hybrid' refers to the fact that the circuit combines discrete diodes, bypass capacitors and other circuit components with inte-

grated-circuit components. The phase shifter of Fig. 6.13 enables the phase shift to be set accurate to within 5 to 8 degrees over a bandwidth

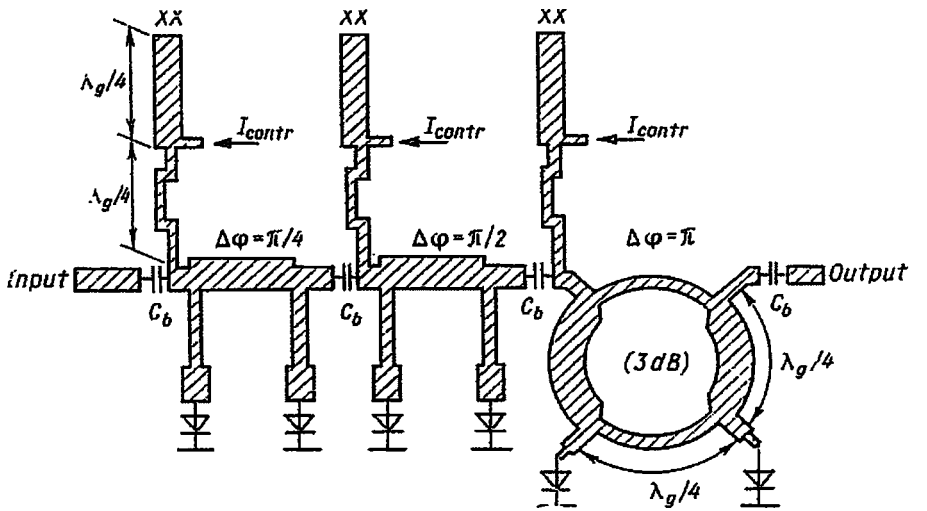


Fig. 6.13. Pattern on the microstrip board of a three-stage microwave switch using PIN diodes (C_b, d.c. blocking capacitor)

of 5 to 10%. It introduces an attenuation of 1.0 to 1.5 dB and has an input SWR of not over 1.3.

6.7. Microwave Ferrite Devices

In electronics, the name 'ferrite' has come to describe practically all magnetic oxides containing iron as the principal component, and is even used sometimes in reference to mixtures of magnetic oxides and certain nonmagnetic oxides.

In general, ferrites have a relative dielectric constant, ε_r, ranging from 5 to 16, and a loss tangent of 10^{-2} or 10^{-3}. They are crystalline materials possessing gyromagnetic properties due to the behaviour of electrons in the atoms that make up the crystal lattice.

The three most common groups of ferrites are those characterized as *spinels*, *garnets*, and *hexagonal ferrites*, so classed according to their crystal structure.

Ferrites having the spinel structure answer the general chemical formula $Me^{II}O \cdot Fe_2O_3$, where Me^{II} represents an ion of a divalent metal (such as Ni, Mn, Cu, Co, Mg, etc.), so there may be nickel ferrite, cobalt ferrite, and so on.

Ferrites having the garnet crystal structure are usually called simply garnets and have the general formula $3Me_2O_3 \cdot 5Fe_2O_3$. Here Me

is more often yttrium, Y, so that there is yttrium-iron garnet (YIG), $3Y_2O_3 \cdot 5Fe_2O_3$.

Ferrites having hexagonal structures actually exist in several families, each characterized by a particular structure but with a variable chemical composition.

Ferrites can be polycrystalline and monocrystalline. The former are fabricated in much the same way as ceramic materials — a mixture of metal oxides and a plasticizer are moulded into green products, and these are then fired at 1000-1400°C.

Monocrystalline (or, more commonly, single-crystal) ferrites are grown by processes similar to those used in the manufacture of semiconductors.

Most often, ferrites come as rods round or rectangular in cross-section, plates, wafers, or polished spheres of a small diameter (around 1 mm) made from single crystals.

Microwave ferrite devices utilize the *gyromagnetic properties* of ferrites. These properties manifest themselves when a ferrite specimen is subjected simultaneously to the action of a d.c. and an r.f. magnetic field. As a result of this exposure, the r.f. magnetic intensity vector $\mathbf{H}_{rf}$ and the r.f. magnetic induction (flux density) vector $\mathbf{B}_{rf}$ are connected in a very complex manner described by the permeability tensor (μ). If the biasing d.c. magnetic field $\mathbf{H}_0$ is oriented along the z-axis of a Cartesian coordinate system, this relation will take the form

$$\mathbf{B}_{rf} = (\mu)\,\mathbf{H}_{rf}, \quad (\mu) = \begin{bmatrix} \dot{\mu} & -j\dot{k} & 0 \\ j\dot{k} & \dot{\mu} & 0 \\ 0 & 0 & \dot{\mu}_0 \end{bmatrix}$$

The tensor components

$$\dot{\mu} = \mu' + j\mu''$$

and

$$\dot{k} = k' + j\dot{k}''$$

are complex quantities, and their dependence on the biasing field intensity (or on the gyromagnetic resonance frequency f_0) has a resonant character (Fig. 6.14*a*). The frequency f_0 is called the *Larmor frequency* and is defined by a well-known relation

$$f_0 = \gamma \mathbf{H}_0$$

where

$$\gamma \approx 3.5 \times 10^{-2} \text{ MHz/(A/m)}$$

is the *gyromagnetic ratio* of an electron, that is, the ratio between its magnetic and mechanical moments. The sharpness of resonance is customarily stated in terms of the gyromagnetic resonance line

width, ΔH, which is the width of the $\mu''(H_0)$ curve at half the maximum level. The value of ΔH can vary between broad limits, depending on the type of ferrite.

Owing to the tensor nature of permeability, an unbounded ferrite medium affects differently right-hand and left-hand circularly polarized electromagnetic waves propagated in the direction of the biasing field. As is known from electrodynamics, for each of these

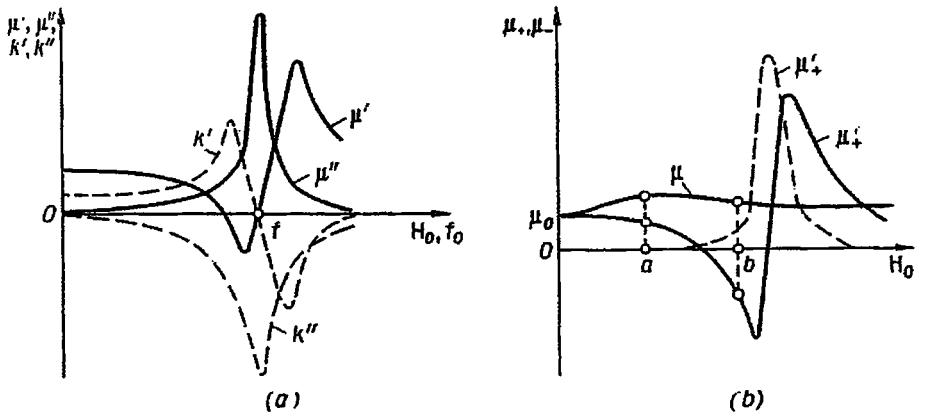


Fig. 6.14. Ferrite parameters as functions of biasing field: *a*, μ-tensor components; *b*, μ_+ and μ_- for rotating polarization fields

waves a ferrite behaves as an isotropic medium whose permeability depends on the direction of rotation of the polarization vector of the wave. For the right-hand polarized wave (the $\mathbf{H}_{rf}$ vector rotates clockwise as seen by an observer looking in the direction of the $\mathbf{H}_0$ field), the permeability

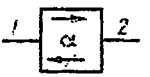
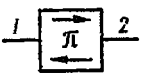
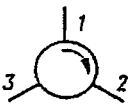
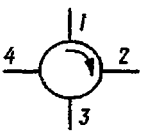
$$\dot{\mu}_+ = \mu'_+ + j\mu''_+$$

shows a resonant behaviour, and the imaginary component μ''_+ which takes care of the losses occurring in the ferrite specimen is maximal at resonance. The resonant behaviour of the function $\dot{\mu}_+(H_0)$ is due to the fact that the direction of rotation for the driving magnetic field is the same as that of the precession of the electronic magnetic moments. For the left-hand polarized wave, the driving magnetic field rotates and the electronic magnetic moments precess in opposite directions, because of which no resonance can take place, and μ_- varies with the biasing field in a quiet manner. The plots of $\dot{\mu}_+(\mathbf{H}_0)$ and $\dot{\mu}_-(\mathbf{H}_0)$ are shown in Fig. 6.14*b*.

Microwave ferrite devices can be classed into two broad groups. The first group covers *non-rectprocal devices*, such as isolators, gyra-

tors, and circulators. Their graphic symbols and ideal scattering matrices are presented in Table 6.1.

Table 6.1

Device	Graphic symbol	Ideal scattering matrix S
Isolator		$\begin{bmatrix} 0 & 0 \\ \exp(-j\varphi) & 0 \end{bmatrix}$
Gyrator		$\begin{bmatrix} 0 & \exp(-j\varphi) \\ \exp(-j\varphi \pm \pi) & 0 \end{bmatrix}$
Three-port circulator		$\exp(-j\varphi)\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$
Four-port circulator		$\exp(-j\varphi)\begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$

A *ferrite isolator* is a two-port device which transmits a wave in one direction almost without reflection or attenuation, but absorbs the wave propagated in the opposite direction. Ferrite isolators are used to protect microwave generators against changes in load impedance, as components of decoupling (isolating) circuits, and also in measurement set-ups.

A *ferrite gyrator* is a non-reciprocal (that is, non-adjustable) phase shifter whose transmission coefficient in one direction of propagation is 180° out of phase with that in the opposite direction. Gyrators are used as building blocks in more complex nonreciprocal devices.

A *circulator* is a match-terminated, lossless nonreciprocal multiport junction in which the microwave power entering port *1* appears at port *2*, whereas the power entering port *2* appears at port *3*, and so on. Most commonly, use is made of three- and four-port circulators, notably in order to permit the use of a common antenna for

transmission and reception. They are also used in parametric amplifiers, power combiners, and similar applications.

The devices in the second group are *control devices*, such as phase shifters, switches, commutators, and tunable filters. Their response is controlled by varying or switching the current in the control windings. Some ferrite devices have *internal magnetic memory* and are controlled by injecting single current pulses into the control windings.

Most microwave ferrite devices are designed for operation in the SHF band. Their use in the EHF band is also possible, but there are some difficulties in making ferrite specimens that would have the desired electric characteristics, and in maintaining control magnetic fields of an appropriate strength.

The principal advantages of ferrite devices are their ability to handle high levels of microwave power and their insensitivity to appreciable short-duration overloads. On the demerit side are the dependence of their performance on temperature and a relatively sluggish response because of the time lag associated with the control magnetic circuit.

6.8. Nonreciprocal and Control Ferrite Devices

Faraday-effect devices. The Faraday effect refers to the rotation of the plane of polarization as a linearly polarized wave is propagated in a gyrotropic medium. It manifests itself when the wave is propagated in the direction of the biasing field.

As will be recalled, a linearly polarized electromagnetic wave can be visualized as the sum of two waves, one positively (clockwise or right-hand) polarized and the other, negatively (counter-clockwise or left-hand) polarized. For each of them, a ferrite specimen acts as an isotropic medium with permeabilities μ_+ and μ_-, respectively. In a biased ferrite specimen the right- and left-hand circularly polarized waves differ in the propagation phase constant

$$\beta_+ = \omega (\varepsilon\mu_+)^{1/2}$$

and

$$\beta_- = \omega (\varepsilon\mu_-)^{1/2}$$

such that $\mu_+ < \mu_-$ and $\beta_+ < \beta_-$ (point *a* in Fig. 6.14*b*). On travelling a distance l along the ferrite specimen, the two waves suffer a different phase shift, $\beta_+ l$ and $\beta_- l$, respectively. As a result, the total linearly polarized E field at the end of the distance l will have been rotated through an angle

$$\theta = (\beta_- - \beta_+)\, l/2$$

in the clockwise direction, as seen looking in the direction of the lines of force of the biasing (d.c.) field $\mathbf{H}_0$. Importantly, the *angle of*

rotation θ *of the plane of polarization is independent of the direction of wave propagation (with or against the* $\mathbf{H}_0$ *vector)*. This fact explains why the Faraday effect is nonreciprocal.

The Faraday effect is widely utilized in microwave isolators and circulators constructed in circular or square guides that transmit waves of any polarization.

A Faraday-rotation isolator (Fig. 6.15*a*) consists of a circular-guide segment enclosing a ferrite rod, *1*, arranged to lie along the

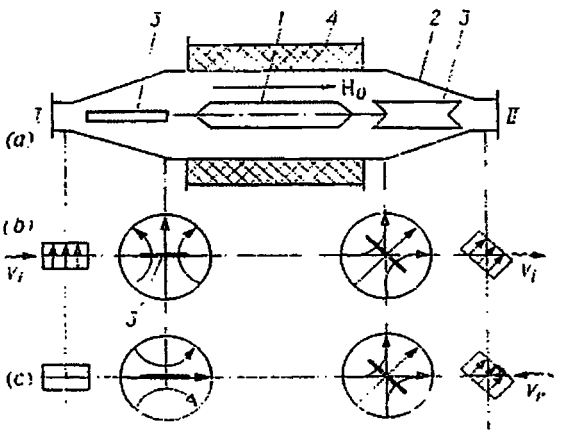


Fig. 6.15. Faraday-rotation isolator in circular guide

guide axis, and an external solenoid, *4*, which sets up a longitudinal biasing field. At both ends, the guide is terminated in tapered circular-to-rectangular guide transitions, *2*. The transitions enclose each an absorbing strip, *3*, set up parallel with the broad faces. The output tapered segment is turned through 45° relative to the input segment. The incident wave v_i entering port *I* is not attenuated by the absorbing strip and is transformed into a TE_{11} wave in the circular guide. The diameter and length of the ferriter rod and the strength of the biasing field are chosen such that, on passing through the ferrite rod, the wave is turned through 45° clockwise and keeps on travelling unattenuated until it enters the output guide segment in which the narrow faces are parallel to the E vector (Fig. 6.15*b*). To minimize reflections, the ends of the ferrite rod and of the absorbing strips are tapered.

The reflected wave v_r entering port *II* (Fig. 6.15*c*) is transformed unattenuated into a TE_{11} mode in the circular guide. As it is propagated through the ferrite rod, its plane of polarization is rotated through 45° clockwise (the direction of Faraday rotation is independent of the direction of propagation and is solely decided by the

direction of the biasing field). At the output from the ferrite rod the E vector is parallel to the broad faces of the rectangular guide at port *1* and the absorbing strip *3*. The reflected wave v_r cannot enter port *1*, and all of the power it carries is dissipated in the absorbing strip. Thus, ideally, this device has the scattering matrix S given in Table 6.1.

The isolator shown in Fig. 6.15 can be converted to a four-port circulator, if we replace the absorbing strips by rectangular guide side tees. The H-planes of these side tees should coincide with the E-planes of the input and output guide segments.

Reversal of the biasing field changes the direction in which the plane of polarization rotates owing to the Faraday effect. Because of this, the sequence of power flow in the forward direction in a Faraday-rotation isolator is reversed. In consequence, reversal of the current which traverses the solenoid transforms an isolator and a circulator to a controlled switch and a commutator.

An important advantage of Faraday-rotation ferrite devices is the fact that they need a relatively weak biasing field (a few thousand amperes per metre in the 3-cm waveband). Among their demerits are the difficulty with which heat can be abstracted from the ferrite rod and a rather bulky size.

Transverse-biased ferrite devices. A rectangular TE_{10} waveguide has two planes parallel to the narrow faces and in these planes the magnetic field due to the travelling wave is circularly polarized. To demonstrate, the components of the magnetic field set up by a TE_{10} mode may be written as

$$\dot{H}_x = j\,(H_A \beta a/\pi) \sin(\pi x/a) \exp(-j\beta z)$$

and

$$\dot{H}_z = H_A \cos(\pi x/a) \exp(-j\beta z)$$

where

$$\beta = 2\pi/\lambda_g$$

is the propagation phase constant.

The components $\dot{H}_x$ and $\dot{H}_z$ are equal in magnitude and the total magnetic field vector H is circularly rotated about the directions parallel to the y-axis in longitudinal planes such that $x = x_0$ and $x = a - x_0$, where

$$x_0 = (a/\pi) \arctan(\lambda_g/2a)$$

In Fig. 6.16*a*, the planes of circular polarization of the H vector are shown by dashed lines. The directions in which the H vector rotates depend on the direction of wave propagation in the guide and are opposite within the segments lying on the opposite sides from the centre line of the guide.

If in a rectangular guide we place along its axis a ferrite strip within the segment sustaining a rotating magnetic field **H** and apply a transverse biasing field $\mathbf{H}_0$ (Fig. 6.16*b*), the ferrite will affect the waves in the guide in the same manner as does an isotropic ferri-

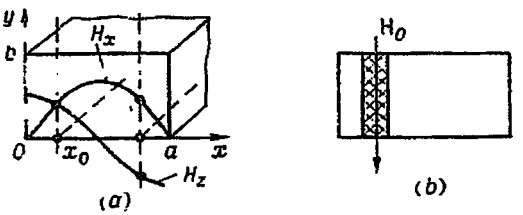


Fig. 6.16. Explaining the effect of magnetic field rotation in rectangular guide: *a*, distribution of H_z and H_x in the guide; *b*, ferrite strip placed in the region of the rotating **H** vector

magnetic specimen whose permeability is different for the waves propagated in the opposite directions (see the plots of μ_+ and μ_- in Fig. 6.14*b*).

The effect of a transverse-biased ferrite strip, *1*, placed in the guide segment supporting a rotating magnetic field **H** is utilized in what

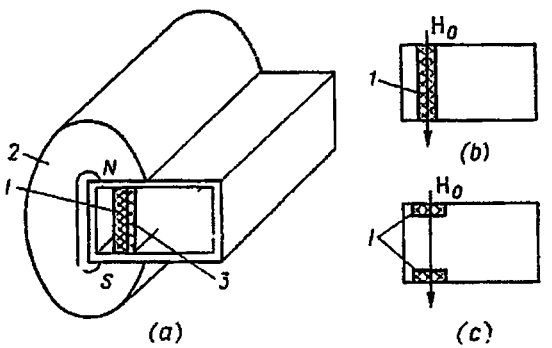


Fig. 6.17. Resonant terminating isolator in rectangular guide

is known as a *resonance isolator* (Fig. 6.17*a*). The transverse biasing field is supplied by a permanent magnet, *2*, and its magnitude is adjusted to be equal to that of the gyromagnetic resonance field for a right-hand circularly polarized wave. As the incident wave passes through the isolator and is propagated in the guide, the ferrite is exposed to the action of the **H** vector which rotates counterclockwise relative to the biasing field, and the wave experiences a relatively small attenuation. During the passage of the reflected wave, the ferrite is acted upon by the clockwise circularly polarized compo-

nent of the H vector, and the wave is heavily attenuated owing to large losses in the ferrite at resonance.

Ferrite strips in resonance isolators can be arranged in any one of two ways, namely in the E-plane (Fig. 6.17*b*) and in the H-plane (Fig. 6.17*c*). In the former case, the permanent magnet may be weaker than in the latter, and the isolator is better matched. The latter arrangement is used in isolators for high-power applications as this improves heat abstraction and minimizes the risk of an electric breakdown.

The biasing field required to produce gyromagnetic resonance and the location of the ferrite strip in the guide depend on frequency —

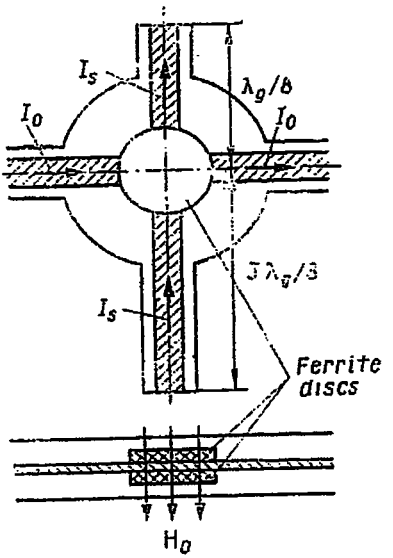


Fig. 6.18. Ferrite isolator in strip line with shorted stubs

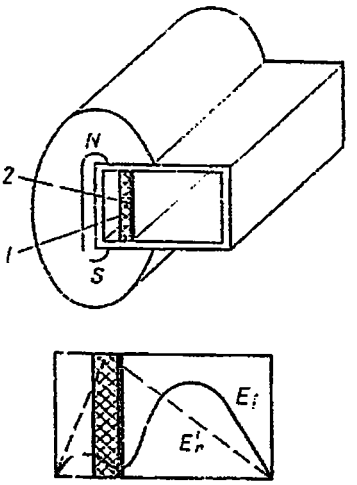


Fig. 6.19. Field-displacement ferrite isolator in rectangular guide

a fact which limits the bandwidth of resonance isolators. The bandwidth of a resonance isolator can be extended by attaching a dielectric slab, *3*, to the ferrite strip. The dielectric slab maintains the conditions for the rotation of the H vector over the bandwidth within the ferrite-loaded section.

Figure 6.18 shows a *stripline resonance isolator*. Here, a rotating magnetic field around the ferrite specimen is produced by two transverse stubs with a length of $\lambda_g/8$ and $3\lambda_g/8$, respectively. The conductors of the stubs carry a current equal in magnitude to that in the main line, but in phase-quadrature with it. The interaction of the currents flowing in the arms of the cross-junction produces a magnetic field which is left-hand circularly polarized with respect to the

biasing field for the incident wave and right-hand circularly polarized for the reflected wave which is absorbed in the ferrite discs at gyromagnetic resonance. The total current branching off into the stubs is zero, so the stubs do not impair the quality of match for the isolator.

A drawback common to resonance isolators is the need to provide a heavy permanent magnet which could supply a magnetic field with an intensity of several hundred thousand amperes per metre in the 3-cm waveband.

Figure 6.19 shows a *field-displacement isolator* in rectangular guide. It has a ferrite strip, *1*, placed in a guide region containing a rotating magnetic field. The transverse biasing field is chosen such that the permeability is negative for the clockwise rotating incident wave (see point *b* in Fig. 6.14*b*). In this case, the propagation constant for the wave in the ferrite is imaginary, the wave cannot be propagated in the ferrite, and the field is crowded out of the specimen. For the left-hand circularly polarized reflected wave, $\mu_- > 0$, and the increased dielectric constant of the ferrite specimen causes the microwave field to concentrate at the ferrite surface. The distributions of the E field for the incident and reflected waves over the cross-section of the ferrite-loaded guide are shown in Fig. 6.19. The ferrite strip, *1*, is given a coat of absorbing metal film, *2*, so the reflected wave concentrating near the ferrite specimen is heavily attenuated. The incident wave remains practically unaffected by the absorbing film. As compared with resonance isolators, the field-displacement type has a lighter magnetic circuit and a wider bandwidth, but can only operate at relatively low power levels.

Phase-shift circulators. Let us refer to Fig. 6.16*b* which shows a ferrite strip placed in a region in a rectangular guide where the total magnetic field is circularly polarized. If the biasing field is other in magnitude than is required for gyromagnetic resonance to occur (say, at point *a* in Fig. 6.14*b*), absorption in the ferrite will be insignificant, but the ferrite will delay the waves propagated in the opposite directions differently. Let the corresponding propagation phase constants be designated as β_+ and β_-. If the ferrite length is l, the phases of the transmission coefficients for the waves propagated in opposite directions and traversing the ferrite-loaded section will differ by an amount called the *nonreciprocal phase shift*

$$\Delta\varphi = (\beta_- - \beta_+)l$$

As often as not, the ferrite length may be chosen such that the nonreciprocal phase shift is equal to 180°, and the ferrite-loaded guide section will then have the properties of a gyrator (see Table 6.1).

When combined with waveguide 3-dB directional couplers, nonreciprocal phase shifters make up what are known as *phase-shift circulators*, arranged as shown in Fig. 6.20*a*. This arrangement contains

two 3-dB directional couplers connected in cascade and phase shifters placed in one or both connecting segments. The nonreciprocal phase shifts should be chosen to suit the type of directional coupler used (quadrature or even-odd-excited) and the direction of circulation.

The circulator shown in Fig. 6.20*b* uses a sidewall single-hole 3-dB coupler, *1*, and a magic-T, *2*. The connecting guide segments which encloses nonreciprocal phase shifters formed by identical ferrite slabs,

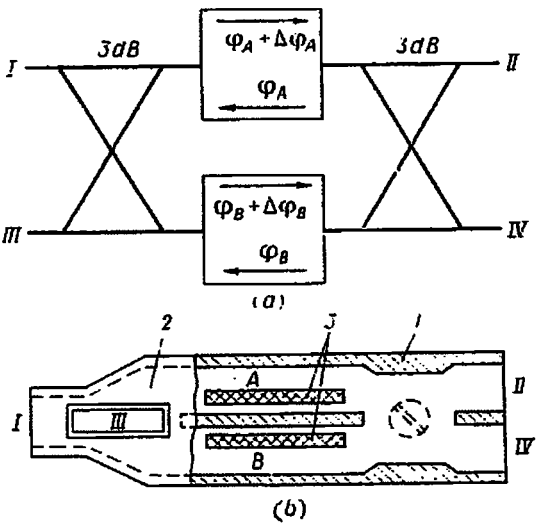


Fig. 6.20. Phase circulator in rectangular guide

3, are placed next to the common narrow face and are transversely magnetized from a common permanent magnet. For waves propagated in opposite directions, the phase shifts produced in the ferrite-loaded guide segments differ by $\pi/2$. For waves propagated in one direction, say, from left to right, the phase shift is $(-\varphi_0 + \pi/2)$, in channel *A* and $-\varphi_0$ in channel *B*. For waves propagated in the opposite direction, the phase shift is $-\varphi_0$ in channel *A* and $-(\varphi_0 + \pi/2)$ in channel *B*. No power transfer can take place between ports *I* and *III* and also between ports *II* and *IV* owing to the decoupling (isolation) properties of the magic-T and the 3-dB directional coupler. The signal entering port *I* is split in the same phase between channels *A* and *B*, and the waves traversing these channels reach port *II* with the same delay equal to $-(\varphi_0 + \pi/2)$ and are combined there. At port *IV*, however, these waves come from port *I* with different phases, $-(\varphi_0 + \pi)$ and $-\varphi_0$, respectively. Thus, they are in anti-phase and cancel each other.

The signal applied to port *II* is split between channels *A* and *B* so that one wave has a phase of zero and the other, a phase of $-\pi/2$. Also, on traversing segments *A* and *B*, they acquire each an excess phase shift of $-\varphi_0$ and $-(\varphi_0 + \pi/2)$, respectively. The result is that they arrive at the ports of the magic-T in anti-phase and are combined in its E-plane tee, that is, at port *III* of the circulator. Signal transfer from port *III* to port *IV* and from port *IV* to port *I* occurs in a similar fashion.

Among the advantages offerred by nonreciprocal phase-shift circulators over Faraday-rotation circulators are a wider bandwidth and the capability to handle high-level power. The latter ability is

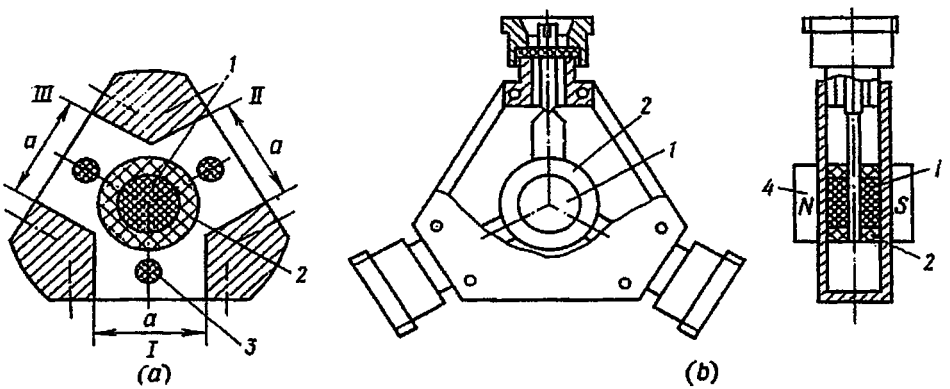


Fig. 6.21. Y-circulators: *a*, waveguide; *b*, stripline

due to the fact that the ferrite slabs cemented to the broad faces of the guides improve heat abstraction. On the demerit side of phase-shift circulators are their increased size and weight because of the need to use two hybrid junctions.

Let us examine more compact three-port circulators. One of them is the *waveguide Y-circulator*. Its key component is an H-plane Y-junction at the centre of which is placed a transversely biased ferrite rod, *1*, in a dielectric sleeve, *2* (Fig. 6.21*a*). The biasing field is supplied by external permanent disc magnets, *4*. Basically, this Y-circulator operates as follows. The wave entering port *I* is split into two waves flowing around the ferrite rod on two sides. For these waves regions of a circularly polarized H vector occur within the ferrite rod. In one region the direction of circular polarization is positive relative to the biasing field, whereas in the other it is negative. Because of the difference in the permeabilities of the ferrite rod, μ_+ and μ_-, the waves flowing round the ferrite rod differ in phase velocity. The sizes and parameters of the ferrite rod are chosen such that the two waves could arrive in phase at port *II* and in anti-phase at

port *III*. Thus, the signal entering port *I* can only be transferred to port *II*. Since Y-circulators possess rotational symmetry, it may be argued that energy can be transferred from port *II* to port *III* and from port *III* to port *I*. The dielectric sleeve enclosing the ferrite rod improves the stability of the device towards variations in temperature and biasing field. The dielectric posts, *3*, ensure broadband matching for the ports of the Y-circulator.

Another form of compact Y-circulators is a *stripline Y-circulator* (Fig. 6.21*b*) which operates in much the same manner as waveguide Y-circulators.

Of all circulators, it is Y-circulators that are most commonly used. They owe their popularity to simple design, small size, and low weight. Also, Y-circulators have a wide broadband. The operating bandwidth of waveguide Y-circulators is 30%, and that of stripline Y-circulators, an octave.

6.9. Ferrite Phase Shifters

Ferrite phase shifters successfully compete with *PIN* diode phase shifters. In the SHF band, they offer a convenient means for handling high levels of power, although their response is more sluggish.

The simplest form is the phase shifter in rectangular guide with an axially biased ferrite rod, known as the *Reggia-Spencer phase shifter*. The ferrite rod, *1*, is positioned along the axis of the rectangular guide and is axially biased by an external control winding, *2* (Fig. 6.22). The ferrite-loaded guide acts as a below-cutoff guide for waves whose E vector is parallel with the broad faces, so the Faraday effect does not take place. The control magnetic field varies the permeability of the ferrite rod and the phase constant of the fundamental mode in the ferrite-loaded guide. The Reggia-Spencer phase shifter is a reciprocal one and can be built for any operating frequency in the range 8-70 GHz. Among its advantages are simple design, the ability to vary phase from zero to 360°, a relatively weak control field, and an insertion loss of as low as 0.5 to 1.0 dB. Unfortunately, the Reggia-Spencer phase shifter can operate over a relatively narrow bandwidth and at a mean power of not over 0.5 kW.

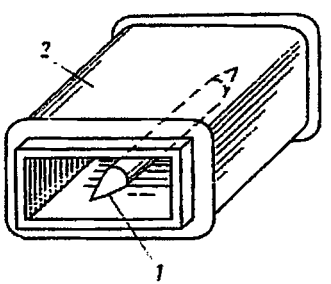


Fig. 6.22. Reggia-Spencer ferrite phase shifter

A wider popularity has been won by phase shifters using *transversely biased ferrite rods*. They utilize the existence of regions sustaining a rotating **H** vector inside the guide and the dependence of the permeabilities of the ferrite rod, μ_+ and μ_-, on the strength

of the biasing field. Therefore, instead of permanent magnets, they use electromagnets.

Transverse-biased ferrite phase shifters are *nonreciprocal* devices because within the ferrite-loaded segment the **H** vector rotates in different senses for the waves propagated in the opposite directions.

As a way of increasing the phase shift and reducing the length,

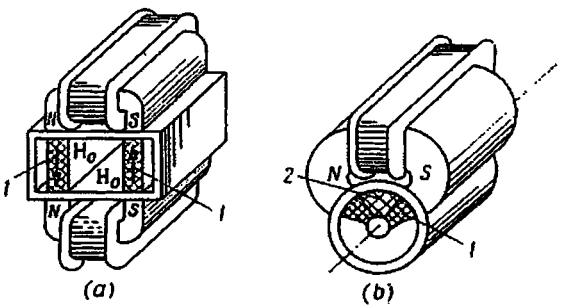


Fig. 6.23. Transverse-biased nonreciprocal ferrite phase shifters: *a*, waveguide; *b*, coaxial

ferrite phase shifters in rectangular guide usually contain two ferrite cards, *1*, positioned on each side of the central plane of the guide and biased in opposite sense (Fig. 6.23*a*).

A ferrite phase shifter can also be constructed in coaxial form (Fig. 6.23*b*). Here, a rotating **H** field is set up in the ferrite specimens by loading part of the cross-section of the transmission line with a dielectric, *2*.

A disadvantage common to all phase shifters with a continuous change of phase (analog phase shifters) is the need to keep the control coils energized all the time. This drawback is circumvented in discrete phase shifters using square *B-H* loop ferrites. Once magnetized, such ferrites will remain in this state for any time, however long, and involve an expenditure of energy only in order to reverse the sense of magnetization. In control devices using square *B-H* loop ferrites it is essential to provide a closed path for the magnetic flux inside the ferrite.

The most commonly used form of the waveguide phase shifter using square *B-H* loop ferrites (the *toroidal phase shifter*) is shown in Fig. 6.24. It consists of several two-state segments which provide between them a set of binary phase states and are controlled by switching their saturation magnetization from one extreme value to the other. A combination of a 180-degree, a 90-degree, and a 45-degree segment will span a phase shift of 0 to 360° in steps of 45 degrees. The ferrites, *1*, are rectangular toroids with their sense of magnetization switched by current pulses injected in opposite directions

through the conductors, *3*, which thread the toroids. Toroidal phase shifters are nonreciprocal devices, and operate on a principle similar to that underlying the action of transverse-biased ferrite-strip phase shifters. At the ends, a toroidal phase shifter has dielectric matching elements, *2*. The switching time of a toroidal phase shifter is from 0.5 to 2.0 μs. Over a bandwidth of ±5% and a phase shift span of

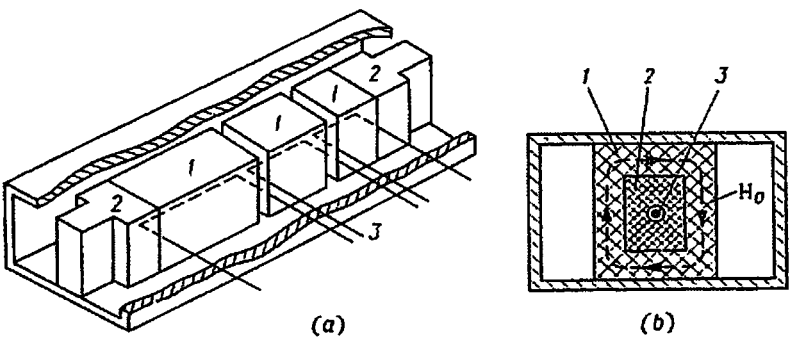


Fig. 6.24. Magnetic-memory toroidal binary phase shifter

0 to 360°, the insertion loss is 0.8-1.2 dB and the input SWR is not over 1.2 in all phase states. The average microwave power rating is 0.2 to 0.4 kW, and the energy required to switch a toroidal phase shifter from one state to another is between 3×10^{-4} and 3×10^{-3} J.

6.10. Magnetically Tunable Ferrite Filters

For their operation, magnetically tunable ferrite filters depend on the ferrimagnetic resonance that takes place in single-crystal ferrite specimens.

The key element of a tunable ferrite filter is a *ferrite* (or *ferrimagnetic*) *resonator* — a magnetized microwave ferrite specimen which has a narrow gyromagnetic resonance line width and is coupled to the electromagnetic field on a transmission line.

As a rule, ferrite resonators are highly polished spheres fabricated from single crystals of yttrium-iron garnet (YIG). Such resonators have an unloaded Q of about 2×10^3 to 3×10^3 and utilize the resonance associated with the right-handed precession of electron-spin magnetic moments. The resonance frequency of a ferrite resonator is

$$f_0 = \gamma H_0$$

where H_0 is the magnetizing (biasing) field strength and $\gamma \approx 3.5 \times \times 10^{-2}$ MHz/(A/m). The resonance frequency is independent of the specimen size, and the resonator can therefore be made very small. The external Q of a ferrite resonator depends on its size and its loca-

tion relative to the line to which it is coupled. Off resonance, a ferrite resonator behaves as an isotropic ferrimagnetic specimen and, being small, has but a negligible effect on the associated line. It is only within a narrow band centred on the frequency of gyromagnetic resonance that the amount of coupling between the ferrite and the line abruptly increases, and the electromagnetic field acquires components non-existent in the original wave structure on the line. Exactly this event is utilized in tunable microwave filters.

Figure 6.25 shows three single-resonator ferrite filters: one using a coupling loop, another built in rectangular guide, and a third constructed in microstrip.

The *loop-coupled filter* (Fig. 6.25*a*) is a ferrite sphere placed at the centre of two intersecting loops lying in mutually perpendicular pla-

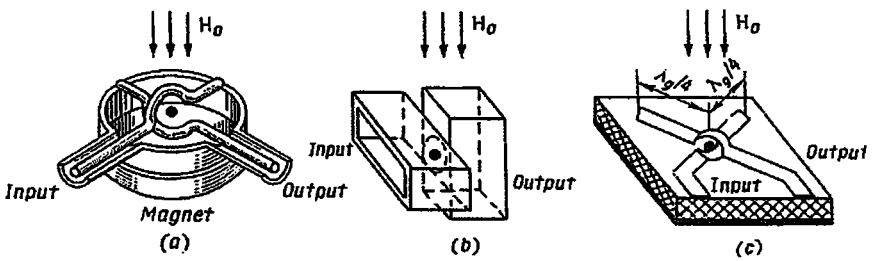


Fig. 6.25. Tuned ferrite-resonator filters

nes. The magnetic fields of the loops are likewise mutually perpendicular, so no energy can be transferred between the loops. At gyromagnetic resonance of the ferrite sphere, the magnetic field, $H_{rf\,x}$, due to one loop acts on the ferrite and gives rise to a magnetic induction, $B_{rf\,y}$, which excites the other loop, so that the signal is free to pass on to the filter output.

The *rectangular-guide filter* (Fig. 6.25*b*) has two in-line rectangular-guide segments turned through 90° relative to each other and coupled by a hole in the common wall. Placed in the centre of the coupling loop is a biased ferrite sphere. Away from gyromagnetic resonance, the guides are decoupled because their dominant modes are orthogonally polarized. At gyromagnetic resonance, the guides are coupled by the off-diagonal components of the permeability tensor of the ferrite, and the microwave signal is free to pass from one guide into the other.

The *microstrip filter* (Fig. 6.25*c*) uses two intersecting microstrips which, in the absence of gyromagnetic resonance, are practically decoupled from each other because there is no coupling via the magnetic field, whereas the electric field at the intersection is negligibly small since the intersection is within a quarter of a wavelength away from the open ends of the microstrips. At the intersection, there

is a ferrite sphere biased by a control field which is at right angles to the base of the microstrip board. At gyromagnetic resonance, the magnetic induction acquires a component oriented along the driving line, and this gives rise to a strong coupling between the microstrips.

The gain-vs-frequency response of the above filters shows well-defined resonance peaks whose location (that is, the resonance frequency) can readily be controlled between broad limits by varying the magnetizing (biasing) field. The shape of the frequency response of tunable filters can markedly be improved using several closely spaced ferrite resonators with a carefully selected amount of coupling between them.

Part Two

Antennas

Chapter Seven

The Electromagnetic Field and Antenna Parameters

7.1. Purpose and Classification of Antennas

Antennas fall in the group of passive components of radio broadcasting, communication, navigation, radar, television and other systems. Physically, an antenna is an assemblage of conductors, dielectrics and ferrites, basically intended to radiate electromagnetic waves into and/or receiving them from space. On top of that, state-of-the-art antennas provide spatial filtration of signals, thereby assuring the required directional characteristics of radio systems, or ensuring the location and ranging of objects in radar applications, or the location of radiation sources.

The performance of antennas is stated in terms of a wide range of radio-engineering, structural, service, and economical parameters and characteristics. The design and the required performance of an antenna depend in a substantial way on the frequency band or bands in which it is to operate. It customary to class antennas into long-wave (or LF), medium-wave (or MF), short-wave (or HF), VHF/UHF, SHF, and optical-wave. *This text is mainly concerned with SHF (or microwave) antennas operating in the range from 300 MHz to 300 GHz* used by the overwhelming majority of radar systems, ground-based and space radio communication, radio telemetry, radio navigation, and television.

Another classification of antennas, based on the manner in which the radiated field pattern is formed, lists four types.

(1) *Small radiators* ($l \leqslant \lambda$) for frequencies in the range 10 kHz-1 GHz. This class includes single-dipole and slot radiators, strip and microstrip antennas, loop antennas, and frequency-independent radiators.

(2) *Travelling-wave antennas* from a wavelength to ten wavelengths in size for frequencies in the range 3 MHz-10 GHz. This class inclu-

des helical, dielectric-rod, Yagi-Uda, impedance, and leaky-wave antennas.

(3) *Antenna arrays* from a wavelength to a hundred or more wavelengths in size for frequencies in the range 3 MHz-30 GHz. Such an antenna consists of a great number of individual radiators. With an independent control of excitation current phase (and, sometimes, current amplitude) for each element of the array, it is possible to control its radiation pattern (or beam) electronically. There may be linear, planar, circular, spherical and conform antenna arrays. (Here, conform means an array conforming to the shape of the object carrying it.) Antenna arrays form a basis for signal-processing antenna systems, including adaptive arrays, that is, antenna systems capable of adaptation to a changing environment.

(4) *Aperture-type antennas* from a wavelength to a thousand wavelengths in size for frequencies in the range 100 MHz-100 GHz and higher. The most common types are reflector, horn and lens antennas. Also related to them are so-called 'hybrid antennas' each of which is a combination of reflectors or lenses and an antenna-array feed. Aperture-type antennas are built on optical principles and provide the highest directivity of radiation.

State-of-the-art antennas are designed by resort to a combination of *theoretical* and *experimental* methods. The trial-and-error approach has to be used because it is difficult, if at all possible, to build sufficiently accurate mathematical models that could adequately represent the radiation pattern and other characteristics of an antenna. Quite often even a very large computer may fail in its resources to carry out rigorous calculations involved in antenna design. Rather, one makes calculations at a reasonable level of mathematical rigour and supplements and combines them with experimental studies. In this combined approach, wide use is made of antenna simulation based on electrodynamic similarity (in which case the physical dimensions of the antenna are changed along with its operating frequency, while its electrical dimensions are kept constant). Simultation markedly speeds up and makes less expensive the experimental studies involved. Recently there has been a growing tendency to use computer-aided antenna measurement ranges: the computer of such a facility measures the antenna performance, processes the results, and presents the data in a format convenient to the designer.

The great multitude of techniques used in antenna analysis and synthesis stems from the need to handle a host of problems some related to systems design, radio engineering and mechanical design, others to automatic control, aerodynamics, meteorological, etc. This turns antenna development and testing into a challenging and creative activity which calls for a wealth of knowledge and, especially, mathematical expertise.

7.2. The Basic Structure of an Antenna

The following elements can be seen in the basic structure of a specific antenna: the input, a matching circuit, a distributor, and a radiating system (Fig. 7.1). The term '*input*' usually refers to a transverse plane on the transmission (feeder) line sustaining the desired mode. The position of this input plane should be stated in exact terms — this is essential for an exact calculation of the transmission system that feeds the antenna. State-of-the-art antennas

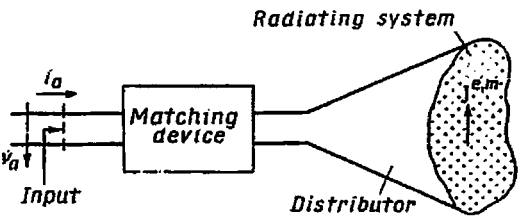


Fig. 7.1. Schematic arrangement of an antenna

may have several (sometimes, several hundred or thousand) inputs. These inputs can be used for simultaneous operation of a single antenna at various frequencies or in order to form several independent and distinct radiation patterns.

The *matching circuit*, or *coupler*, has as its objective to ensure that the feeding line operates under conditions as close to a travelling wave as practicable. Apart from the usual circuits providing narrowband and broadband matching, antennas often use input matching through the proper choice of some structural dimensions for the distributor.

The *distributor* of an antenna is an assemblage of conductors and dielectrics and has as its objective to distribute the radiating currents in a fashion that would provide the desired directional properties.

Finally, the *radiating system* of an antenna is a region in space traversed by the currents that excite electromagnetic waves. In view of the reciprocity principle the same name can be retained for receiving antennas as well. The radiating system may be either a set of real currents flowing over metal surfaces or fictitious equivalent electric and magnetic currents on the closed surfaces surrounding the antenna [see Eqs. (A.14) and (A.15) in the Appendix] or electric and magnetic polarization currents in the volumes taken up by dielectric and ferrimagnetic materials.

The separation of the distributor from the radiating system within the antenna structure has its explanation in the fact that antenna analysis and synthesis is traditionally decomposed into two problems: an internal problem and an external problem. The *internal*

problem has to do with finding the r.f. current distribution functions in the radiating system. The *external problem* is concerned with determining the electromagnetic field of the antenna and its parameters (beam width, side lobe level, directivity, and the like) from the known current distribution.

The separation of antenna analysis and synthesis into the internal and external problems is warranted in two cases, namely: (1) when the objective is to develop approximate methods for the *analysis* of an antenna of known design, based on guessing the likely

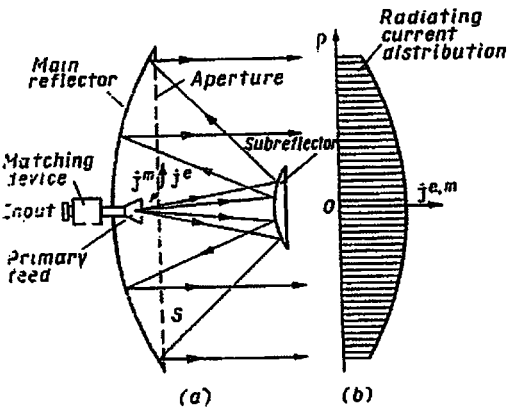


Fig. 7.2. Two-reflector parabolic antenna: *a*, general arrangement and ray paths; *b*, distribution of radiating currents over radius

solution for the more elaborate internal problem, and (2) when the objective is to develop methods for the *synthesis* of an antenna with a radiation field having the desired characteristics. In such a case, the preliminary determination of the current distribution in the radiating system facilitates the design of the respective distributor.

Example 7.1. A two-reflector parabolic antenna (Fig. 7.2). The radiating system of this antenna is its aperture, that is, an imaginary plane surface covering the exit opening of the main parabolic reflector. It is assumed that this surface is traversed by equivalent electric and magnetic currents. The main reflector and the subreflector form, together with a primary horn feed, the distributor structure. A separate matching device is located where the input waveguide changes into the horn feed.

The aperture size of a parabolic antenna is tens or even hundreds of times the operating wavelength. Therefore, the distributor can be designed by invoking the techniques of geometrical optics (the rays in such antenna travel the same paths as in an optical projector). Sometimes, when a more accurate design of a parabolic antenna is desired, the real electric currents induced by the field of the subre-

flector and the primary feed on the surface of the main reflector are used as the radiating system instead of the equivalent electric and magnetic currents in the aperture. Thus, the radiating system of one and the same antenna can be determined differently, depending on the approach used.

7.3. Calculation of the Electromagnetic Fields Set Up by Radiating Systems in the Far, Transition and Near Zones

In theoretical studies, the radiating system of an antenna is usually assumed to be located in an unbounded, homogeneous space. This assumption materially simplifies the calculation of the electromagnetic fields and provides an opportunity to take into account the effect of the Earth and surrounding objects by resort to the methods of diffraction theory at a later time.

As electrodynamics tells us, the vector potentials of the electromagnetic field set up by a known distribution of exciting electric and magnetic currents $\dot{\mathbf{J}}^{e,m}(x', y', z')$ at an arbitrary observation point $P(x, y, z)$ are given by

$$\dot{\mathbf{A}}^{e,m}(x, y, z) = \frac{1}{4\pi} \int_V \dot{\mathbf{J}}^{e,m}(x', y', z') \frac{\exp(-j\beta r)}{r} dV' \qquad (7.1)$$

where

$$r = [(x - x')^2 + (y - y')^2 + (z - z')^2]^{1/2}$$

is the distance between the observation point P and the integration point Q, $\beta = 2\pi/\lambda$, and V is the volume taken up by the currents of the radiating system.

Equation (7.1) is the exact solution of some inhomogeneous Helmholtz equations [see Eqs. (A.4) and (A.5) in the Appendix]. This is the unique solution because it satisfies the condition of radiation at infinity and has a finite value everywhere. Substitution of (7.1) into (A.4) offers an opportunity to determine the field vectors $\dot{\mathbf{E}}$ and $\dot{\mathbf{H}}$ at any point in space.

We may write in shorthand

$$\begin{aligned} \dot{\mathbf{E}}(x, y, z) &= \hat{\mathscr{E}}\{\dot{\mathbf{J}}^{e,m}(x', y', z')\} \\ \dot{\mathbf{H}}(x, y, z) &= \hat{\mathscr{H}}\{\dot{\mathbf{J}}^{e,m}(x', y', z')\} \end{aligned} \qquad (7.2)$$

where $\hat{\mathscr{E}}\{\cdot\}$ and $\hat{\mathscr{H}}\{\cdot\}$ are integrodifferential vector operators specifying the sequence of calculation of the desired field components. These operators establish correspondence between the specified di-

stribution of electric and magnetic currents in the volume V and the distribution of the $\dot{\mathbf{E}}$ and $\dot{\mathbf{H}}$ fields in space. They are rigorous and applicable for any likely relative positions of source points and observation points. With this approach, however, it is usually not possible to obtain simple closed expressions even for relatively simple ra-

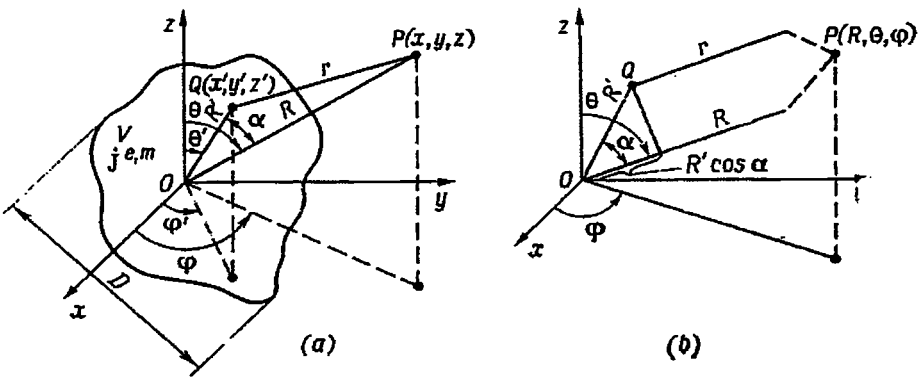


Fig. 7.3. To calculation of the electromagnetic fields set up by a radiating system: *a*, the general case; *b*, observation point in the far-field region

diating systems. This is the reason why one has to make simplifying assumptions which involve breaking up space into the *far*, *transition* and *near zones*.

Let us introduce a spherical coordinate system R, θ, φ, with its centre located inside the radiating system (Fig. 7.3*a*). Let also the points Q (x', y', z') and P (x, y, z) represent, respectively, the present integration point inside the radiating system and the observation point in the surrounding homogeneous medium. The distance r in Eq. (7.1) is

$$r = QP = (R^2 + R'^2 - 2RR' \cos\alpha)^{1/2}$$

where α is the angle between the directions OQ and OP.

If $R > R'$ and the observation point is sufficiently remote from the volume enclosing the radiating currents, the distance r may approximately be written as a series in the powers of the ratio:

$$r = R\,[1 - (R'/R)\cos\alpha + (R'^2/2R^2)\,(1 - \cos^2\alpha) + (R'^3/2R^3)\cos\alpha\,(1 - \cos^2\alpha) + \ldots] \qquad (7.3)$$

Let $R \gg R'$, which corresponds to the *far-zone* (or *radiation*) *field* most important to antenna theory. (It is also called the *far field* or *Fraunhofer region*). Then Eq. (7.1) simplifies as follows:

(1) In the denominator of the integrand we may approximately set $r = R$, in which case the term $1/R$ can be moved outside the integral sign.

(2) In the exponent of the exponential function in the integrand we may set

$$r \approx R - R' \cos \alpha$$

in which case the function $\exp(-j\beta r)$ may also be moved outside the integral sign.

The more delicate approach to the replacement of r by an approximate expression in the exponent of the exponential function is due to the fact that here the discarded terms must be small in comparison with 2π which is the period of the exponential function having an imaginary exponent. Actually, the second assumption means that the rays drawn to the observation point in the far-zone field from the origin of coordinates and from the present integration point Q are deemed parallel (see Fig. 7.3*b*). The excess term $R'\cos\alpha$ added to the distance r is called the *path length difference*. The path length difference takes into account the relative delay of the spherical waves arriving at the observation point from two elementary sources, one located at the origin of coordinates and the other at the point $Q(x', y', z')$.

The path length difference $R'\cos\alpha$ is in effect a projection (see Fig. 7.3*b*) of the vector

$$\mathbf{R}' = \mathbf{i}_x x' + \mathbf{i}_y y' + \mathbf{i}_z z'$$

on the direction of the unit vector drawn from the origin of coordinates to the observation point

$$\mathbf{R}/R = \mathbf{i}_x \sin\theta\cos\varphi + \mathbf{i}_y \sin\theta\sin\varphi + \mathbf{i}_z \cos\theta$$

On multiplying together these vectors in scalar form, we obtain an explicit expression for the path length difference

$$R' \cos\alpha = x' \sin\theta\cos\varphi + y' \sin\theta\sin\varphi + z' \cos\theta \qquad (7.4)$$

Using the simplifying assumptions introduced in Eq. (7.1), we obtain an asymptotic equation for the vector potential in the far-zone field

$$\dot{\mathbf{A}}_{\infty}^{e,m}(R, \theta, \varphi) = \frac{\exp(-j\beta R)}{4\pi R} \int_V \dot{\mathbf{j}}^{e,m}(x', y', z') \exp(j\beta R' \cos\alpha)\, dV' \qquad (7.5)$$

Here, the subscript '∞' indicates that this expression holds for R tending to infinity. The limit of applicability for (7.5) will be defined a bit later.

As follows from (7.4), the value of the integral in (7.5) solely depends on the angular coordinates of the observation point and is independent of the distance R. To pass from the vector potentials $\dot{\mathbf{A}}_{\infty}^{e,m}$ to the $\dot{\mathbf{E}}$ and $\dot{\mathbf{H}}$ field vectors, we should perform the operation of spatial differentiation prescribed by Eqs. (A.4). On performing

several identity transformations and discarding the terms showing a radical dependence of the form $1/R^2$ (or $1/R^3$), that is, immaterial in the far-zone field, we obtain the following design equations:

$$\left.\begin{aligned} \dot{E}_\theta &= (-j2\pi/\lambda)\,(Z_m \dot{A}^e_{\theta\infty} + \dot{A}^m_{\varphi\infty}), \quad \dot{H}_\varphi = \dot{E}_\theta/Z_m \\ \dot{E}_\varphi &= (-j2\pi/\lambda)\,(Z_m \dot{A}^e_{\varphi\infty} - \dot{A}^m_{\theta\infty}), \quad \dot{H}_\theta = -\dot{E}_\varphi/Z_m \end{aligned}\right\} \tag{7.6}$$

where

$$Z_m = (\mu_a/\varepsilon_a)^{1/2}$$

is the characteristic wave or, simply, wave impedance of the medium.

In practical calculations, it is convenient to evaluate integrals of the form (7.5) in terms of Cartesian components

$$\dot{A}^{e,\,m}_{x,\,y,\,z,\,\infty} = \frac{\exp(-j\beta R)}{4\pi R} \int\limits_V \dot{J}^{e,\,m}_{x,\,y,\,z} \exp[j\beta (x' \sin\theta\cos\varphi + y' \sin\theta \sin\varphi + z'\cos\theta)] \times dx'\,dy'\,dz' \tag{7.7}$$

and to pass to spherical coordinates with the aid of the following relations:

$$\begin{aligned} \dot{A}_\theta &= \dot{A}_x \cos\theta\cos\varphi + \dot{A}_y \cos\theta \sin\varphi - \dot{A}_z \sin\theta \\ \dot{A}_\varphi &= -\dot{A}_x \sin\varphi + \dot{A}_y \cos\varphi \end{aligned} \tag{7.8}$$

Let us formulate the principal properties of the electromagnetic field set up by the radiating system in the far zone:

1. The far-zone field is a transverse one, which means that the $\dot{\mathbf{E}}$ and $\dot{\mathbf{H}}$ vectors have no components in the direction of wave propagation.

2. The far-zone field in the vicinity of the observation point is a plane electromagnetic wave, which means that the $\dot{E}_\theta$ and $\dot{H}_\varphi$ components as well as the $\dot{E}_\varphi$ and $\dot{H}_\theta$ components are respectively in phase, and their ratio is equal to the wave impedance of the medium.

3. The dependence of the field on the distance R has the form of a diverging spherical wave, $\exp(-j\beta R)/R$. However, equi-phase surfaces for each field component are not, in the general case, spheres with centre at the origin of coordinates because $\dot{E}_\theta$ and $\dot{E}_\varphi$ are complex functions of the angles θ and φ, whereas the origin of coordinates is chosen arbitrarily.

4. The angular distribution of the $\dot{\mathbf{E}}$ vector components in the far-zone field is independent of the distance R and may be described by

the following functions:

$$\dot{F}_\theta(\theta, \varphi) = \frac{\dot{E}_\theta(\theta, \varphi)}{|\dot{E}_{\theta\max}(\theta_1, \varphi_1)|}$$

$$\dot{F}_\varphi(\theta, \varphi) = \frac{\dot{E}_\varphi(\theta, \varphi)}{|\dot{E}_{\varphi\max}(\theta_2, \varphi_2)|}$$

where θ_1, φ_1 and θ_2, φ_2 are the directions of maximum radiation for the respective components. The functions $\dot{F}_\theta(\theta, \varphi)$ and $\dot{F}_\varphi(\theta, \varphi)$ describe what are called the *normalized radiation patterns* for the respective components.

Sometimes, the property in (4) is used in an 'inverted' form, in which case the observation points referred to the far-zone field are those for which the angular behaviour of the transverse components is independent of the distance to the antenna.

5. The radiation power flow in the far-zone field is always radial. The power flux density is equal to the radial component of the Poynting vector

$$\Pi_R = 0.5\mathrm{Re}\exp(\dot{E}_\theta \dot{H}_\varphi^* - \dot{E}_\varphi \dot{H}_\theta^*)$$

Since

$$\dot{H}_\varphi = \dot{E}_\theta / Z_m \text{ and } \dot{H}_\theta = -\dot{E}_\varphi / Z_m$$

we get

$$\Pi_R(\theta, \varphi) = (|\dot{E}_\theta(\theta, \varphi)|^2 + |\dot{E}_\varphi(\theta, \varphi)|^2)/2Z_m$$

The imaginary part of the Poynting vector in the far-zone field is zero. Thus, the power flux density in each direction is defined as the sum of the independent power flux densities associated with the meridional and azimuthal components of the field.

The angular function

$$F^2(\theta, \varphi) = \Pi_R(\theta, \varphi)/\Pi_{R\max}(\theta_0, \varphi_0) \qquad (7.9)$$

where $\Pi_{R\max}$ is the magnitude of the Poynting vector in the direction of maximum radiation, θ_0, φ_0, describes what is called the *normalized power radiation pattern.*

Let us see at what distance from the radiating system it is legitimate to use Eqs. (7.5) and (7.6) for field calculations. In other words, let us define the boundary of the far-zone field. The basic assumption that we have used consists in the replacement of the exact expression

$$r = (R^2 + R'^2 - 2RR'\cos\alpha)^{1/2}$$

by an approximate one,

$$r \approx R - R'\cos\alpha$$

The phase error arising because of this in the exponent in the integrand of Eq. (7.1) is, in view of the expansion in (7.3), approxima-

tely equal to

$$(\beta R'^2 \sin^2 \alpha)/2R$$

Since the maximum value of R' is equal to about half the maximum dimension D of the radiating system (Fig. 7.3), it follows that the maximum phase error may be equal to $\beta D^2/8R$. Assuming that the allowable phase error is $\pi/8$ (or 22.5°) and carrying out simple manipulations, we find that the sought-for distance to the near boundary of the far zone is $R > 2D^2/\lambda$. The boundary of the far zone rapidly rolls back as the size of the radiating system in wavelengths is increased. At $D/\lambda = 10$, the far zone begins at a distance $R > 200\lambda$, and at $D/\lambda = 100$, the far zone begins at a distance of $20\,000\lambda$.

Since the radiation pattern of an antenna defines the angular behaviour of radiation fields precisely in the far zone, it is essentially important to satisfy the condition $R > 2D^2/\lambda$ when the radiation pattern is plotted experimentally by moving a receiving antenna probe on a spherical surface around the antenna under test.

At $R < 2D^2/\lambda$, the far-zone field gradually changes into the *transition-zone field* (or the Fresnel region). In the calculating the transition-zone field of a radiating system, the following simplifying assumptions are made.

1. As with the far-zone field, the quantity r in the denominator of the integrand in (7.1) is taken to be approximately equal to R and is placed outside the integral sign.

2. In the exponent of the integrand of (7.1) it is assumed that

$$r \approx R - R' \cos \alpha + R'^2 (1 - \cos^2 \alpha)/2R$$

which is equivalent to discarding all terms raised to higher than the second power in the power series (7.3). The function $\exp(-j\beta R)$ which is independent of the coordinates of the sources, is placed outside the integral sign.

Thus, in the transition zone the vector potentials are given by

$$\dot{A}^{e,m}_{Fr}(R, \theta, \varphi) = \frac{\exp(-j\beta R)}{4\pi R}$$

$$\times \int_V \mathbf{J}^{e,m}(x', y', z') \exp\{j\beta[R' \cos\alpha - R'^2(1 - \cos^2\alpha)/2R]\}\, dV' \quad (7.10)$$

where the path length difference $R' \cos \alpha$ is given by Eq. (7.4) as before.

3. In carrying out spatial differentiation according to Eqs. A.4, one discards all terms showing a radial dependence of the form $1/R^2$ and $1/R^3$, similarly to how this was done in calculating the far-zone field. In consequence, the components of the $\dot{\mathbf{E}}$ and $\dot{\mathbf{H}}$ field vectors

in the transition zone can be found by Eqs. (7.6) on replacing the vector potentials $\dot{A}_{\infty}^{e,m}$ by the vector potentials $\dot{A}_{Fr}^{e,m}$.

The conclusions formulated earlier as regards the transverse nature of the far-zone field and its local similarity with a plane electromagnetic wave in the vicinity of any observation point remain valid for the transition zone as well. However, the dependence of the field on distance no longer has the character of a spherical wave, $\exp(-j\beta R)/R$, because the distance R is additionally included in the exponent of the integrand of Eq. (7.10). Calculations show that this is responsible for the fact that a damped oscillation superimposed on the transition-zone field monotonically diminishes as $1/R$. The angular distribution of the field vector components turns out to be dependent on the distance R, which means that the radiation pattern of the radiating system in the transition zone is distorted progressively more as R decreases in value. The analysis which has as its objective to determine the accuracy of the adopted approximation for the transition zone shows that the distance should lie between the limits $D/4 + (D/2)\ (D/\lambda)^{1/3} \leqslant R \leqslant 2D^2/\lambda$, where D is the maximum dimension of the radiating system.

The term $D/4$ on the left-hand side of the above inequality is of importance solely for very small antennas and takes care of the error that arises because $1/r$ is replaced by $1/R$ and because $1/R$ is then placed outside the integral sign in (7.1). For $D = 10\lambda$, the transition zone extends between the limits $13.5\lambda \leqslant R \leqslant 200\lambda$. As the antenna increases in size, the transition zone expands so that at $D = 100\lambda$ it extends in size from 250λ to $20\,000\lambda$. A more rigorous analysis would show that the boundaries of the transition and far zones of a radiating system depend not only on the distance R, but also on the angle subtended from the observation point. These boundaries are furthermore dependent on the shape of the radiating system and the form of the current distribution $\dot{J}^{e,m}(x', y', z')$.

Located at distances shorter than the near boundary of the transition zone is the near zone of the radiating system. The near-zone field has a complex configuration, and its calculation calls for the use of the rigorous operators defined in (7.2). In the general case, the near-zone field contains all the field components and depends on the distance R in an irregular manner, whereas the Poynting vector becomes complex and may run in a direction other than the radius vector **R**.

The near-zone field stores some electromagnetic energy — a fact which handicaps obtaining a good match for the antenna input over a wide frequency range.

7.4. The Complex Vector Radiation Pattern of an Antenna

By analogy with the field set up by the Hertzian dipole [see Eqs. (A.12a) in the Appendix], the far-zone electromagnetic field due to an arbitrary antenna may be written as

$$\dot{E} = (Z_m/2\lambda)\, \dot{I}^e_A h_e \dot{F}(\theta, \varphi) \exp(-j\beta R)/R \qquad (7.11)$$

where $\dot{I}^e_A$ = complex amplitude of current at the selected point A of the radiating system (as a rule, it lies at the distribution maximum or the input)

$Z_m = (\mu_a/\varepsilon_a)^{1/2}$ = characteristic wave impedance of the medium

λ = wavelength in the medium

h_e = effective length of the antenna

$\dot{F}(\theta, \varphi)$ = complex vector radiation pattern function of the antenna

The complex vector radiation pattern function $\dot{F}(\theta, \varphi)$ appearing in Eq. (7.11) characterizes the angular distribution of the field and its polarization and phase properties. In specifying this vital parameter of an antenna, it is mandatory to state the location of the origin of coordinates relative to which the phase difference is reckoned.

In the most general case, the function $\dot{F}(\theta, \varphi)$ is the product of three factors

$$\dot{F}(\theta, \varphi) = \dot{F}(\theta, \varphi)\, \dot{p}(\theta, \varphi) \exp[j\Phi(\theta, \varphi)] \qquad (7.12)$$

describing, respectively, the magnitude, polarization and phase structure of the far-zone field of the antenna. Consider each factor separately and in turn.

The field-strength radiation pattern $F(\theta, \varphi)$. This real, positive function is normalized so that $\max F(\theta, \varphi) = 1$. Obviously, the square of this function, $F^2(\theta, \varphi)$, coincides with the normalized angular distribution of the Poynting vector, as defined by Eq. (7.9).

The field-strength radiation pattern of an antenna can be obtained by both calculation and experiment. It is customary to visualize it as a plot of some shape. Any graphical representation is based on the fact that the field-strength radiation pattern is conceived as a surface joining the tips of the vectors radiating from the origin of coordinates and having, in each direction, a length equal on the scale used to the value of the function $F(\theta, \varphi)$. The most common forms are toroidal, pencil-beam, fan-beam and cosecant beam radiation patterns.

A distinction of the *toroidal radiation pattern* (Fig. 7.4*a*) is the fact that the radiation is distributed nearly uniformly in a plane at

right angles to the axis of the toroid. Toroidal-beam antennas are mainly used in radio communication, radio navigation, and radio broadcasting. The *pencil-beam radiation pattern* shows a nearly symmetrical main lobe against the background of a multiplicity of side lobes (Fig. 7.4*b*). In the *fan-beam radiation pattern*, the main beam markedly differs in width in two mutually perpendicular planes (Fig. 7.4*c*). Pencil-beam and fan-beam antennas are mostly used in radar and communication radio systems. In the *cosecant radiation*

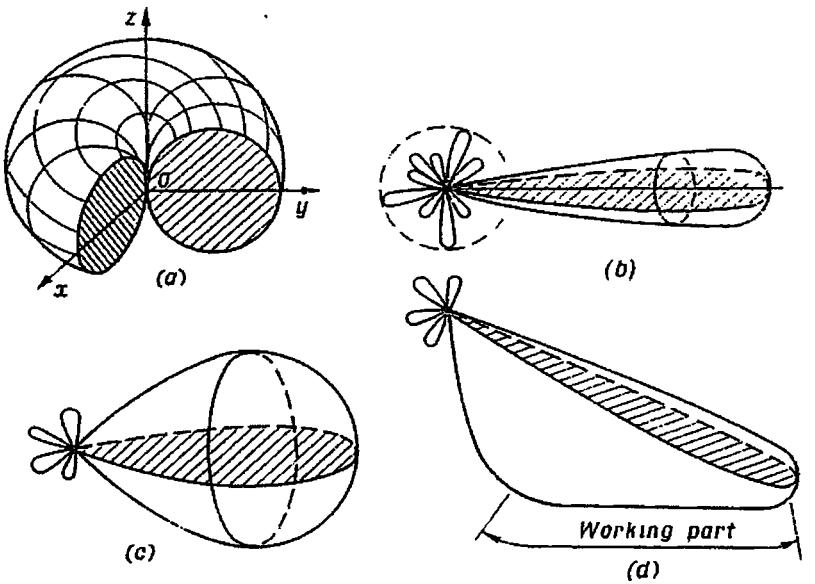


Fig. 7.4. Radiation patterns

pattern the main lobe is unsymmetrical in shape (Fig. 7.4*d*), its operating part in one of the planes (usually, vertical) being defined as $F(\theta) = \operatorname{cosec} \theta$, while in the other plane the mean lobe is symmetrical and small in width. Cosecant-beam antennas are preferable in air-borne ground-surveillance and ground-based air-surveillance radars. The operating portion of the cosecant beam ensures that the echo signals have nearly the same strength on any slant range to a target.

A three-dimensional representation of the function $F(\theta, \varphi)$ such as shown in Fig. 7.4 is difficult to construct and is not informative enough. That is the reason why it is usual to assess the shape of a 3-dimensional radiation pattern from its sections at some selected planes. For antennas of low directional properties, use is made of the principal sections of a spherical system of coordinates: the equatorial plane and a pair of meridional planes. For highly directional

pencil-beam and fan-beam antennas, it is more common to use pairs of perpendicular sections passing through the direction of maximum radiation. One of the sections is chosen in the plane where the main lobe has a minimum width. For linearly polarized antennas, use may also be made of a pair of sections parallel to the E and H vectors.

Radiation pattern sections may be constructed in polar or Cartesian coordinates on various scales, namely: linear (when the mag-

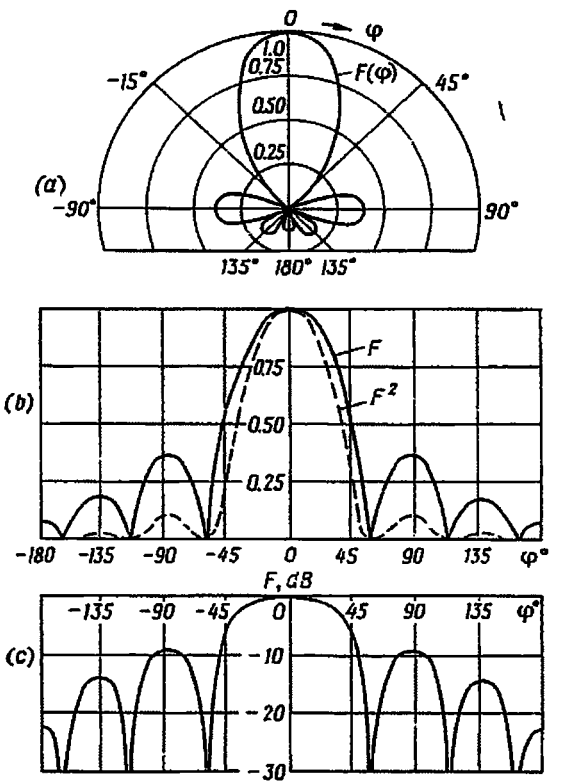


Fig. 7.5. Graphical representation of two-dimensional radiation patterns: *a*, field-strength radiation pattern in polar coordinates; *b*, field-strength and power radiation pattern in Cartesian coordinates; *c*, logarithmic radiation pattern in Cartesian coordinates

nitude of the field is plotted), quadratic (when the power of the field is plotted), or logarithmic (the decibel scale). Several forms of the same two-dimensional radiation pattern are shown in Fig. 7.5. Polar diagrams are easy to grasp, but the angular positions of radiation maxima and minima are difficult to determine. The quadratic scale

tends to mask the side lobes and is therefore unable to show the radiation pattern of antennas with a low side-lobe level. The logarithmic scale is introduced by the relation

$$F_{\text{dB}}(\theta, \varphi) = 20 \log F(\theta, \varphi) = 10 \log F^2(\theta, \varphi)$$

and gives a good idea about field-strength radiation patterns over a wide dynamic range.

Since the advent of computer graphics, there has been a growing trend to use cartographic methods for the presentation of three-dimensional radiation patterns. This purpose is usually served by a

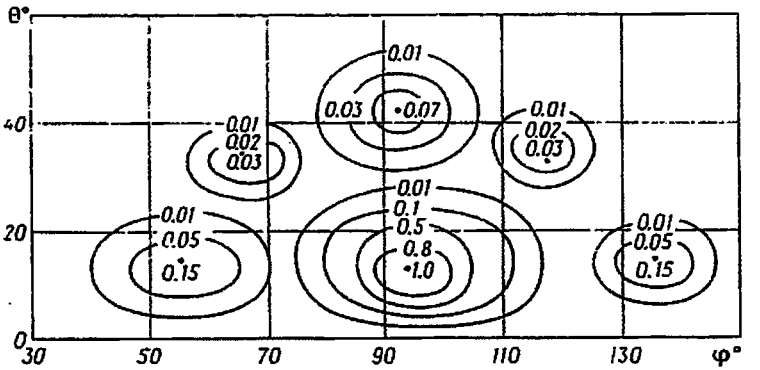


Fig. 7.6. Cartographic representation of a radiation pattern

suitable angular coordinate grid on which are superimposed closed contours of the function $F(\theta, \varphi)$ at different levels. The various levels are marked by distinct colours or numbers. The simplest equidistant projection uses a square coordinate grid in the directions θ and φ (Fig. 7.6). This projection is convenient in depicting the main lobes and some of the side lobes in the neighbourhood.

Polarization properties. The vector factor $\dot{\mathbf{p}}(\theta, \varphi)$ in (7.12) is a unit polarization vector with two components oriented along the basis unit vectors of a spherical coordinate system

$$\dot{\mathbf{p}}(\theta, \varphi) = \mathbf{i}_\theta \dot{p}_\theta(\theta, \varphi) + \mathbf{i}_\varphi \dot{p}_\varphi(\theta, \varphi) = (\mathbf{i}_\theta, \mathbf{i}_\varphi) \begin{bmatrix} \dot{p}_\theta \\ \dot{p}_\varphi \end{bmatrix} = (\mathbf{i}\dot{\mathbf{p}}) \qquad (7.13)$$

The vector $\dot{\mathbf{p}}$ has a magnitude of unity irrespective of direction, that is,

$$|\dot{p}_\theta|^2 + |\dot{p}_\varphi|^2 = 1$$

The components $\dot{p}_\theta$ and $\dot{p}_\varphi$ give an idea about the relative size of the vertical and horizontal components of the far-zone field due to

the antenna in a selected direction and also the phase shift between them.

In the general case, the two components of the polarization vector are complex numbers. One of them is usually deemed to be real (which means that the phase of this component is included in the imaginary exponent in the third factor of (7.12)) and is symbolized as α. This is the *principal component of polarization*, always stipulated in specifications for the design of an antenna.

The other component of the polarization vector, orthogonal to the principal one, may be called the *spurious* (or *cross-polarization*) *component of polarization*. On recalling the notation for the principal polarization component, we may write

$$\dot{\mathbf{p}}(\theta, \varphi) = \mathbf{i}_p \alpha(\theta, \varphi) + \mathbf{i}_s (1 - \alpha^2)^{1/2} \exp[j\psi(\theta, \varphi)] \qquad (7.14)$$

where $\mathbf{i}_p$ = basis unit vector of principal polarization
$\alpha(\theta, \varphi)$ = real positive function
$\mathbf{i}_s$ = basis unit vector of spurious (cross) polarization
$\psi(\theta, \varphi)$ = phase shift between the components

The quantity $\alpha^2 \leqslant 1$ is the polarization efficiency of an antenna and shows the fraction of the total power flux density radiated in the principal polarization in a given direction. The term $1 - \alpha^2$ is the fraction of the power flux density due to cross polarization.

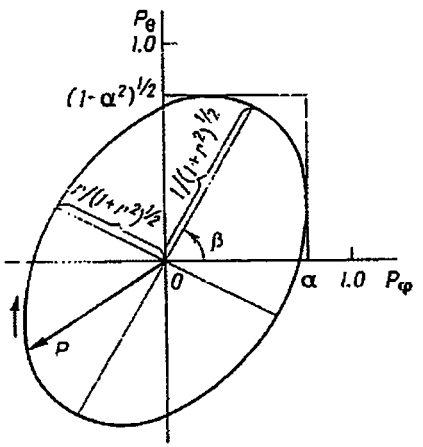


Fig. 7.7. Polarization ellipse

Let us look into the behaviour of the instantaneous value of the total polarization vector. Figure 7.7 shows a plane tangent to the spherical front of the radiated wave in the vicinity of the observation point (the wave deproaches from the observer beyond the plane of the drawing). Suppose that the component in the φ direction corresponds to the principal polarization. Then the instantaneous values of the projections of the total polarization vector will be

$$p_\varphi = \alpha \sin \omega t$$

and

$$p_\theta = (1 - \alpha^2)^{1/2} \sin(\omega t + \psi)$$

The locus of the positions taken up by the tip of the polarization vector at consecutive instants of time is then an *ellipse*. Inside this ellipse, the total polarization vector (and, in consequence, the vector $\dot{\mathbf{E}}$) rotates in a periodical fashion so that a complete rotation occupies a time span equal to the period of oscillation of the carrier whereas

the direction of rotation depends on the sign of the phase angle, $-\pi \leqslant \psi \leqslant \pi$. For positive values of ψ, the rotation is clockwise (right-hand rotation); for negative values of ψ, the rotation is counter-clockwise (left-hand rotation). Thus, the polarization vector rotates in the direction of the component which is lagging in phase—the observer looks in the wake of the deproaching wave.

Quantitatively, the polarization ellipse is characterized in terms of (1) the ratio of the minor to the major axis, $r \leqslant 1$ and (2) the angle of orientation of the major axis, β (Fig. 7.7). The axial ratio is assigned a "+" sign when $\dot{\mathbf{p}}$ rotates clockwise and a "—" sign when $\dot{\mathbf{p}}$ rotates counter-clockwise.

Construction of the polarization basis. Note that the polarization vector will remain unaltered if between the two matrix factors on the right-hand side of (7.13) we place a second-order unit matrix written as the product of two unitary conjugate factors set up in the fashion of the canonical matrix defined in (3.18):

$$\dot{\mathbf{p}} = (\mathbf{i}_\theta, \mathbf{i}_\varphi)\, \mathbf{U}\mathbf{U}_t^* \begin{bmatrix} \dot{p}_\theta \\ \dot{p}_\varphi \end{bmatrix} = (\mathbf{i}_p, \mathbf{i}_c) \begin{bmatrix} \dot{p}_p \\ \dot{p}_c \end{bmatrix}$$

where

$$(\mathbf{i}_p, \mathbf{i}_c) = (\mathbf{i}_\theta, \mathbf{i}_\varphi)\, \mathbf{U}$$

and

$$\begin{bmatrix} \dot{p}_p \\ \dot{p}_c \end{bmatrix} = \mathbf{U}_t^* \begin{bmatrix} \dot{p}_\theta \\ \dot{p}_\varphi \end{bmatrix}$$

Thus, we have resolved the polarization vector on a new basis whose unit vectors are

$$\mathbf{i}_p = \mathbf{i}_\theta \cos\tau + \mathbf{i}_\varphi \sin\tau \exp(j\psi_1)$$

and (7.15)

$$\mathbf{i}_c = \mathbf{i}_\theta \sin\tau \exp(j\psi_2) - \mathbf{i}_\varphi \cos\tau \exp[j(\psi_1 + \psi_2)]$$

where the parameters τ, ψ_1 and ψ_2 define the unitary matrix $\mathbf{U}$.

Each unit vector of the polarization basis in (7.15) has a magnitude of unity and describes a common-mode elliptically polarized wave. A distinction of such waves is *orthogonality* in the sense that the scalar product of $\mathbf{i}_p$ and $\mathbf{i}_c$ reduces to zero:

$$\langle \mathbf{i}_p \mathbf{i}_c \rangle = 0$$

The orthogonality of the vectors $\mathbf{i}_p$ and $\mathbf{i}_c$ implies that they contribute to the transfer of power radiated by an antenna independently of each other.

In practice, use is often made of a polarization basis which consists of two circular polarizations opposite in the sense of rotation. The respective parameters of the unitary matrix are $\tau = \pi/4$, $\psi_1 = -\pi/2$,

$\psi_2 = \pi/2$, $\mathbf{i}_p = 0.707\ (\mathbf{i}_\theta - j\mathbf{i}_\varphi)$, and $\mathbf{i}_c = 0.707\ (j\mathbf{i}_\theta - \mathbf{i}_\varphi)$. The principal polarization corresponds to right-hand rotation, and the cross-polarization to left-hand rotation. In the most general case, with arbitrary parameters τ, ψ_1 and ψ_2, the vectors $\mathbf{i}_p$ and $\mathbf{i}_c$ have the same axial ratio, but the major axes of the ellipses at every point in space are at right angles to each other, and the senses of rotation are opposite.

Thus, in order to completely describe the polarization properties of the far-zone field of an antenna, it suffices to specify the required basis and to obtain the functional relationships defining the polarization efficiency $\alpha^2(\theta, \varphi)$ and the phase shift between the principal and cross polarization components, $\psi(\theta, \varphi)$.

For a joint description of the field-strength and polarization properties of an antenna, resort may be made to the radiation pattern function for the particular polarization of the field

$$F_p(\theta, \varphi) = \frac{F(\theta, \varphi)\,\alpha(\theta, \varphi)}{|F(\theta, \varphi)\,\alpha(\theta, \varphi)|_{\max}}$$

$$F_c(\theta, \varphi) = \frac{F(\theta, \varphi)\,[1-\alpha^2(\theta, \varphi)]^{1/2}}{|F(\theta, \varphi)\,[1-\alpha^2(\theta, \varphi)]^{1/2}|_{\max}}$$

It is this type of radiation pattern that is measured in experimental studies of specific antennas.

The phase characteristic of an antenna. The imaginary exponent $\Phi(\theta, \varphi)$ in the third factor of (7.12) is referred to as the *directional phase characteristic* of an antenna for the principal polarization. The function $\Phi(\theta, \varphi)$ characterizes variations in the phase shift of the principal polarization component as the observation point is moved over the surface of a large sphere of radius R with centre at the origin of the chosen coordinate system and therefore materially depends on its choice.

The phase characteristic $\Phi(\theta, \varphi)$ is supplemented by equiphase surfaces in the far-zone field, that is, surfaces on which the phase of the principal polarization component is the same for any angle of observation. In view of the radial dependence of the far-field phase, $-2\pi R/\lambda$, the equation describing an equiphase surface may be written as

$$R(\theta, \varphi) = R_0 + (\lambda/2\pi)\,\Phi(\theta, \varphi)$$

If an equiphase surface is a sphere (minus the likely abrupt changes by $\lambda/2$ as the field-strength radiation pattern function crosses zero), its centre is called the *phase centre* of an antenna. For a remote observer, the phase centre is that point of the antenna from which the spherical waves of the radiated field start out.

The simplest phase characteristic of an antenna is an unvarying function

$$\Phi(\theta, \varphi) = \Phi_0 \pm \pi$$

where Φ_0 is a constant. In such a case, equiphase surfaces are spheres, and the phase centre occurs at the origin of coordinates. If, on the other hand, the function $\Phi(\theta, \varphi)$ is other than unvarying, any one of two cases are possible, namely: (1) the antenna has a phase centre which does not occur at the origin of coordinates, or (2) the antenna has no phase centre at all. In each case, there is a way of simplifying the form of the phase characteristic by shifting the origin of coordinates as appropriate.

Let us refer to Fig. 7.8 which shows the position of the origin of a new coordinate system, point O', whose coordinates in the old coordinate system were x_0, y_0, z_0. In the new coordinate system R', θ, φ, the initial phase characteristic changes owing to the path length difference $R_0 \cos \alpha$ and will have the form [see (7.4)]

$$\Phi'(\theta, \varphi) = \Phi(\theta, \varphi) - \beta R_0 \cos \alpha = \Phi(\theta, \varphi) - \beta (x_0 \sin \theta \cos \varphi + y_0 \sin \theta \sin \varphi + z_0 \cos \theta). \quad (7.16)$$

If an antenna does have a phase centre (case 1), the coordinates x_0, y_0, z_0 can be selected so that $\Phi'(\theta, \varphi)$ is a constant. This can only happen if the initial phase characteristic has the form

$$\Phi_0(\theta, \varphi) = \beta (x_0 \sin \theta \cos \varphi + y_0 \sin \theta \sin \varphi + z_0 \cos \theta + \rho)$$

where ρ is a constant.

It may therefore be argued that an antenna will have a phase centre only if its phase characteristic is representable as the function $\Phi_0(\theta, \varphi)$. This condition was established by A.R. Volpert of the Soviet Union in 1961. In practice, many antennas (horn, helical, turnstile, etc.) do not have a phase centre in a strict sense. But even for them it is possible to indicate a point (the so-called *radiation centre*) relative to which the equiphase surface differs least of all from a spherical surface, and the phase characteristic is constant very nearly.

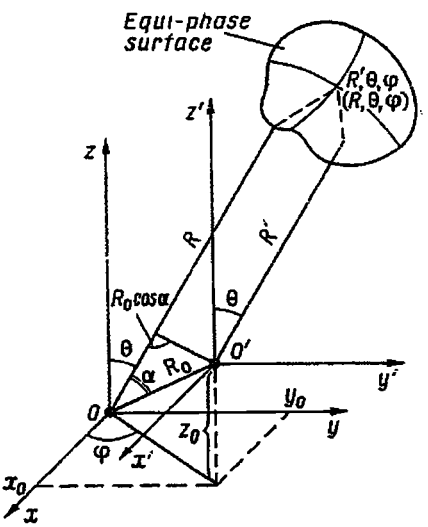


Fig. 7.8. To determining the phase characteristic of an antenna

It is to be stressed that the concept of the phase centre of an antenna and the concept of the radiation centre apply to the principal-polarization component of radiation. For the cross-polarized field,

the directional phase characteristic may be found by the relation

$$\Phi_c(\theta, \varphi) = \Phi(\theta, \varphi) + \psi(\theta, \varphi)$$

where $\Phi(\theta, \varphi)$ = phase characteristic for the principal polarization
$\psi(\theta, \varphi)$ = phase shift between the cross-polarization component and the principal polarization component

7.5. Secondary Directional Parameters of an Antenna

Directivity of an antenna. The extent to which the radiation of an antenna is concentrated in the direction of the main lobe of the field-strength radiation pattern is stated in terms of directivity. This parameter introduced into antenna theory by A.A. Pistolkors of the Soviet Union in 1929 is defined as the ratio of the magnitude of the Poynting vector in the direction of maximum radiation at a distance R in the far-zone field to the average magnitude of the Poynting vector at the surface of the sphere of the same radius, enclosing the antenna

$$D_{max} = \Pi_{max}/\Pi_{av} \text{ for } R = \text{const} \to \infty \qquad (7.17)$$

There is another definition of the directivity of an antenna as a parameter which shows how many times the radiated power should be increased if a real directional antenna be replaced by a hypothetical absolutely nondirectional, isotropic antenna so that the magnitude of the Poynting vector at the observation point will remain unchanged. This definition stresses the fact that an increase in the directivity of an antenna is equivalent, in a way, to an increase in transmitter power. Note that an isotropic antenna for which the power radiation pattern function $F^2(\theta, \varphi)$ is identically equal to unity and the polarization of radiation remains constant is physically unrealizable. It has been proved that the radiation field of a real antenna either has a direction of zero radiation or the polarization substantially depends on direction, and the axial ratio, r, can take on any value between -1 and $+1$.

The power radiated by a directional antenna is proportional to the integral

$$P_\Sigma = A \oint_{\Omega=4\pi} F^2(\theta, \varphi)\, d\Omega, \quad d\Omega = \sin\theta\, d\theta\, d\varphi$$

where A is a proportionality factor. The integral is taken over a complete solid angle $\Omega = 4\pi$, that is, from zero to 2π with respect to φ, and from zero to π with respect to θ. If the antenna is a hypothetical isotropic radiator for which $F^2(\theta, \varphi)$ is identically equal to unity, then its radiated power is given by

$$P_{\Sigma iso} = A \oint_{\Omega=4\pi} 1 \times d\Omega = 4\pi A$$

Using the second definition of the directivity of an antenna, we obtain for the direction of maximum radiation where $F^2_{max}(\theta_0, \varphi_0) = 1$,

$$D_{max} = P_{\Sigma iso}/P_{\Sigma} = 4\pi \Big/ \oint\limits_{\Omega=4\pi} F^2(\theta, \varphi)\, d\Omega \qquad (7.18)$$

In our further discussion we will always mean the directivity of an antenna in the direction of maximum radiation, but the subscript 'max' will be dropped for brevity. For other directions of radiation, the directivity of an antenna is given by

$$D(\theta, \varphi) = DF^2(\theta, \varphi)$$

Example 7.2. For a Hertz dipole in free space, the normalized radiation pattern function has the form

$$F^2(\theta, \varphi) = \sin^2\theta$$

Substituting this function in Eq. (7.18) gives $D = 1.5$. If, however, the dipole is set up at zero height at right angles to an infinite, ideally conducting plane, then $D = 3$, although the normalized radiation pattern function remains the same. The doubling of the directivity is due to the fact integration with respect to θ in the denominator of (7.18) is limited to the interval $(0, \pi/2)$. The point is that now the dipole radiates in a half-space, and the Poynting vector there is quadrupled in magnitude owing to the fact that there appears an image of the dipole [see Fig. (A.4) in the Appendix].

Beamwidth and sidelobe level. Apart from the directivity D, the directional properties of an antenna are assessed in terms of the angular width of the main beam in the radiation pattern in some plane at some specified relative power level. This width is called the *beamwidth* of the antenna in the specified plane, symbolized as $\Delta\theta$ in this text. Most often, one is concerned with the *half-power beamwidth*, that is, the angular width measured between reference lines where the radiation field is 0.707 of the maximum intensity of the main lobe, or 3 dB down from it. Use is also made, but far more seldom, of the beamwidth measured between reference lines of zero radiation intensity; it is symbolized as $\Delta\theta_0$ and is termed the *zero-power beamwidth.*

Still another parameter of no less importance is the *sidelobe level* of an antenna. Most often, the maximum of the largest sidelobe relative to the maximum intensity of the main lobe is stated. When the radiated field has a complex polarization structure, the sidelobe level is stated for both the primary and the cross polarization components.

The beamwidth and sidelobe level of an antenna are parameters which define the resolving power and noise immunity of radio (broadcasting, radar, navigation and other) systems. This explains

why these parameters figure prominently in the specifications for antennas. Their values are verified when an antenna is turned over for use and checked at regular intervals in service.

The directivity of an antenna as a function of beamwidth and sidelobe level. For pencil-beam antennas, the integral in the denominator of Eq. (7.18) may be written as a sum:

$$I_\Sigma = \oint\limits_{\Omega=4\pi} F^2\,d\Omega = \int\limits_{\Omega_m} F^2\,d\Omega + \int\limits_{4\pi-\Omega_m} F^2\,d\Omega = I_m + I_s$$

where Ω_m is the solid angle occupied by the main lobe of the radiation pattern (as determined between lines of zero radiation intensity).

The first terms, I_m, is proportional to the radiation power carried by the main lobe, whereas I_s is proportional to the radiation power carried by the side and backward lobes of the radiation pattern. Now we can separate the expression for the directivity of an antenna, Eq. (7.18), into two factors as follows:

$$\begin{aligned} D &= (4\pi/I_m)\,(I_m/I_\Sigma) = (4\pi/I_m)\,[(I_\Sigma - I_s)/I_\Sigma] \\ &= D'\,(1 - \beta_s) \end{aligned}$$

The first factor is the *directivity of the antenna in terms of the main lobe of the radiation pattern*

$$D' = 4\pi/I_m = 4\pi \Big/ \int\limits_{\Omega_m} F^2(\theta, \varphi)\,d\Omega \tag{7.19}$$

Such would be the directivity of a hypothetical antenna with a single main lobe and with no side or backward lobes.

The quantity $\beta_s < 1$ in the second factor of the expression for D is called the *scatter ratio* defined as

$$\beta_s = I_s/I_\Sigma = \int\limits_{4\pi-\Omega_m} F^2(\theta, \varphi)\,d\Omega \Big/ \oint\limits_{\Omega=4\pi} F^2(\theta, \varphi)\,d\Omega$$

and shows what fraction of the power radiated by an antenna is carried by the side and backward lobes of the radiation pattern. Accordingly, the quantity $1 - \beta_s$ is the fraction of the radiated power concentrated in the main lobe of the radiation pattern. In other words, it defines the *main lobe efficiency* of the antenna.

Example 7.3. Let us approximate a real pencil-beam radiation pattern by resort to a simplified function (Fig. 7.9) equal to unity within a small solid angle

$$\Omega_m = \Delta\theta\Delta\varphi$$

(this is the "main lobe") and having a small value $t < 1$ elsewhere (this is the "effective sidelobe level"). Using this function in the de-

finition of the directivity of an antenna, Eq. (7.18), we find

$$D = 4\pi/[\Delta\theta\Delta\varphi + t^2 (4\pi - \Delta\theta\Delta\varphi)]$$

For not too narrow beams and for low sidelobe levels, that is, on satisfying the inequality

$$\Delta\theta\Delta\varphi \gg t^2 (4\pi - \Delta\theta\Delta\varphi)$$

the main beam efficiency is close to unity, and the directivity of the antenna is practically the same as the directivity of the main beam

$$D' = 4\pi/(\Delta\theta\Delta\varphi) \approx 40\,000/(\Delta\theta\Delta\varphi)$$

As the main beam becomes narrower, however, while the sidelobe level remains unchanged, the scatter ratio increases in magnitude. When

$$\Delta\theta\Delta\varphi = t^2 (4\pi - \Delta\theta\Delta\varphi) \approx 4\pi t^2$$

the main beam efficiency is 0.5. That is, the directivity of the antenna is halved in comparison with D'. When the main beam width is

$$\Delta\theta = \Delta\varphi = 1^\circ$$

the directivity of the antenna is halved at a sidelobe level equal to $t = 0.005$, or at —46 dB. As the beam becomes progressively narrower, with t remaining at a constant value, the scatter ratio tends to unity, the main beam efficiency approaches zero, and the directivity of the antenna tends to a constant value given by

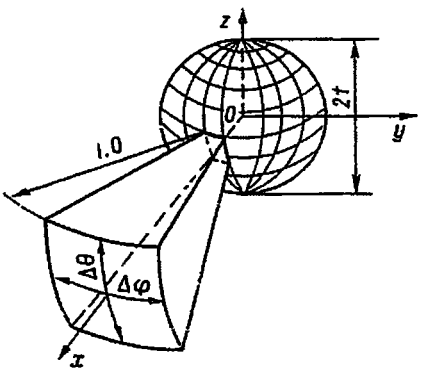


Fig. 7.9. Idealized radiation pattern of an antenna

$$\lim \{D\} = 1/t^2 \text{ for } \Delta\theta\Delta\varphi \to 0 \text{ and } t = \text{const}$$

Thus, when the sidelobe radiation is uniform and has an effective value of t (in terms of field strength), the directivity of the antenna is limited to

$$D_{\text{lim}} = 1/t^2$$

with any beams, however narrow.

In most real antennas, the sidelobe level tends to decrease very rapidly on moving away from the main lobe of the radiation pattern while the main beam width is usually not so small for the main beam efficiency to fall below 0.8. For such antennas, the directivity is commonly estimated for engineering purposes by the equation

$$D = 32\,000/(\Delta\theta\Delta\varphi) \qquad (7.20)$$

where $\Delta\theta$ and $\Delta\varphi$ are the half-power widths (in degrees) of the main lobe of a real antenna. Indeed, it can be shown by simple calculations that when the main lobe of the radiation pattern has a smooth shape the power it carries is approximately equal to the radiation power of an idealized sectoral lobe at the same maximum field strength and with a width in each of two orthogonal planes equal to the respective half-power beamwidth of the real radiation pattern.

7.6. The Transmitting Antenna as a Two-Port

We set out to derive relations for the mathematical modelling of a transmitting antenna from its input to the observation point in the far zone. Before we can do this, we need to refine and interpret several more antenna parameters, such as radiation resistance, effective (or radiating) length, input impedance, efficiency, and gain.

The radiation resistance and effective (radiating) length of an antenna can be connected to the already defined directivity of the antenna as follows. Using the Poynting vector method [see Eqs. (A.13) and the comments], compute the radiated power that sets up the electromagnetic field defined in (7.11):

$$P_\Sigma = \frac{1}{2} |\dot{I}_A|^2 R_\Sigma = \oint\limits_{\Omega=4\pi} \frac{|\dot{E}|^2}{2Z_m} R^2 \, d\Omega = \frac{Z_m h_e^2}{8\lambda^2} |\dot{I}_A|^2 \oint\limits_{\Omega=4\pi} F^2(\theta, \varphi) \, d\Omega$$

The definite integral in the above expression is, according to Eq. (7.18), equal to $4\pi/D$, where D is the directivity of the antenna in the direction of maximum radiation. After obvious cancellations, we obtain the *radiation resistance* referred to point A

$$R_\Sigma = (Z_m \pi / D) (h_e/\lambda)^2$$

and the *effective (radiating) length* of the antenna (likewise referred to the specified point A of the radiating system)

$$h_e = \lambda (D R_\Sigma / Z_m \pi)^{1/2} \tag{7.21}$$

It may be thought that the effective length of an antenna establishes a formal connection between the directivity and radiation resistance of the antenna, when the wavelength and the parameters of the medium surrounding the antenna are known. Note that for the Hertz dipole the effective length is the same as that of the dipole itself.

Now let us substitute the expression for the effective length of an antenna defined by Eq. (7.21) into Eq. (7.11) defining the electric field intensity and re-arrange the factors in some way. This gives

$$\frac{\dot{E}}{(2Z_m)^{1/2}} = \dot{I}_A (R_\Sigma D / 8\pi)^{1/2} \dot{F}(\theta, \varphi) \exp(-j\beta R)/R \tag{7.22}$$

The vector $\dot{\vec{\mathscr{E}}} = \dot{\mathbf{E}}/(2Z_m)^{1/2}$ on the left-hand side of Eq. (7.22) has the dimensions of $(\mathrm{W/m})^{1/2}$ and is the same as the vector $\dot{\mathbf{E}}$ in direction and phase. The square of the magnitude of $\dot{\vec{\mathscr{E}}}$ is exactly equal to the magnitude of the Poynting vector, therefore we will, in our further discussion, call $\dot{\vec{\mathscr{E}}}$ the *radiation intensity vector*.

The factor $\dot{I}_A(R_2/2)^{1/2}$ on the right-hand side of (7.22) can be written in terms of the radiated power of the antenna as

$$(P_\Sigma)^{1/2} \exp(j\varphi_A)$$

where the phase φ_A is the same as that of the radiating current $\dot{I}_A$. In view of the new notation, Eq. (7.22) may be recast as

$$\dot{\vec{\mathscr{E}}}(R, \theta, \varphi) = \exp(-j\varphi_A)(P_\Sigma D/4\pi)^{1/2} \dot{\mathbf{F}}(\theta, \varphi) \exp(-j\beta R)/R \qquad (7.23)$$

where D is the total directivity of the antenna in the direction of maximum radiation, and the complex vector radiation pattern satisfies the normalization to a unit maximum. A limitation of Eq. (7.23) is the fact that it does not take into account the inevitable loss of some of the power applied to the antenna dissipated as heat in non-ideal conductors and insulators. On introducing the *efficiency* of an antenna

$$\eta = P_\Sigma/P_{in} = P_\Sigma/(P_\Sigma + P_{loss})$$

where P_{in} is the total input power, P_Σ is the radiated power and P_{loss} is the ohmic (dissipative) loss in the antenna and the coupler [see also Eq. (A.14)], we derive from (7.23) the following relation:

$$\dot{\vec{\mathscr{E}}}(R, \theta, \varphi) = \exp(j\varphi_A) P_{in}^{1/2} (D\eta/4\pi)^{1/2} \dot{\mathbf{F}}(\theta, \varphi) \exp(-j\beta R)/R \qquad (7.24)$$

The product of the directivity by the efficiency of an antenna, appearing in (7.24), is called its *gain*

$$G = D\eta$$

The gain of an antenna shows how many times the input power should be increased when a real, lossy directional antenna is replaced by a hypothetical, absolutely nondirectional, lossless antenna, while retaining the magnitude of the Poynting vector at the observation point unchanged. Resort to antenna gain as a parameter in its own right is related to the fact that this quantity lends itself readily to a direct measurement by the *comparison method*. In this method, one uses what is called a standard antenna, that is, one whose gain G_s is precisely known and which produces a certain radiation intensity level, $|\dot{\vec{\mathscr{E}}}_s|$, in the far-zone field in operation. The standard antenna is then replaced by the antenna under test, and its input power is adjusted until the radiation intensity at the obser-

vation point is again equal to $|\dot{\vec{\mathcal{E}}}_s|$. Now the gain of the antenna under test can be found as the product of the known gain of the standard antenna by the ratio of input powers obtaining in the former and the latter cases (as a rule, this ratio is determined by a calibrated attenuator). If, however, there is a need to know separately the directivity and the efficiency of an antenna, one should additionally measure the field-strength radiation pattern of the antenna under test, find the directivity by resort to Eq. (7.18), and, finally, determine the efficiency on dividing the gain by the directivity.

Let us go back to Eq. (7.24) and establish the connection between the input power P_{in} and the conditions obtaining at the antenna input. At the input, we choose a fixed transverse plane cutting the incoming transmission line and located ahead of the coupler in the single-mode region (see Fig. 7.1). The normalized incident and reflected voltage waves $\dot{v}_i$ and $\dot{v}_r$ or the total normalized voltage $\dot{v}$ and current i connected to $\dot{v}_i$ and $\dot{v}_r$ by the relations

$$\dot{v} = \dot{v}_i + \dot{v}_r = \dot{v}_i (1 + \dot{\rho})$$

and

$$\dot{i} = \dot{v}_i - \dot{v}_r = \dot{v}_i (1 - \dot{\rho})$$

(where $\dot{\rho}$ is the input reflection coefficient) determine the conditions obtaining at the antenna input. Recall that the magnitude of the normalized incident voltage wave is defined as the square root of the transmitted power (with the dimensions of $W^{1/2}$), and the phase of the normalized incident voltage wave $\dot{v}_i$ is the same as that of the transverse electric field due to the incident wave on the input line.

The ratio of the total normalized voltage and current at the input of an antenna defines the *normalized input impedance* of the antenna

$$\dot{z} = r + jx = \dot{v}/\dot{i} = (1 + \dot{\rho})/(1 - \dot{\rho})$$

or the *normalized input admittance* of the antenna*

$$\dot{y} = g + jb = 1/\dot{z} = (1 - \dot{\rho})/(1 + \dot{\rho})$$

In terms of normalized voltages and currents, the input power of an antenna may be written as

$$P_{in} = \mathrm{Re}\,(\dot{v}\dot{i}^*) = |\dot{i}|^2 r = |\dot{v}|^2 g = |\dot{v}_i|^2 (1 - |\dot{\rho}|^2)$$

---

* For antennas with a TEM-mode feed lines, it is likewise legitimate to use non-normalized input impedance and admittance.

Likewise, several alternative expressions exist for the radiation intensity vector of an antenna:

$$\dot{\mathscr{E}}(R, \theta, \varphi) = \left\{ \begin{array}{l} |\dot{i}|\, r^{1/2} \\ |\dot{v}|\, g^{1/2} \\ |\dot{v}_i|\, (1 - |\dot{\rho}|^2)^{1/2} \end{array} \right\} e^{j\varphi_A} \left(\frac{G}{4\pi}\right)^{1/2} \dot{\mathbf{F}}(\theta, \varphi) \frac{e^{-j\beta R}}{R} \quad \begin{array}{l} (7.25a) \\ (7.25b) \\ (7.25c) \end{array}$$

Equations (7.25) give a normalized description of the 'transit' action of any transmitting antenna from input to the observation point in the far-zone field. These equations are convenient in that they do not contain in an explicit form either antenna dimensions or wavelength. It is enough to know the normalized radiation pattern vector $\dot{\mathbf{F}}(\theta, \varphi)$, antenna gain G, and normalized input impedance (or admittance, or reflection coefficient).

One more remark is in order as regards Eq. (7.25c). The quantity $(1 - |\dot{\rho}|^2)$ in this equation clearly shows how the input power of an antenna is reduced owing to a mismatch at its input. Also, in addition to this term, Eq. (7.25c) contains antenna gain, G. Therefore, it is sometimes practised, when using Eq. (7.25c), to include the quantity $(1 - |\dot{\rho}|^2)$ in the equivalent antenna gain

$$G_{eq} = G(1 - |\dot{\rho}|^2) = D\eta(1 - |\dot{\rho}|^2)$$

which, in addition to the ohmic losses in the antenna, takes care of the power lost due to reflections from its input.

7.7. The Bandwidth and Maximum Operating Power of an Antenna

The parameters introduced in the previous sections characterize an antenna in the case of monochromatic oscillations, that is, in operation at some particular frequency. Any real radio system, however, uses a spectrum of frequencies and provision has often to be made for a change from one frequency to another. Therefore, the key parameters of an antenna include its *bandwidth* — the range of frequencies over which the other of its parameters will not go beyond the tolerances stated in the specifications.

As a rule, the bandwidth of an antenna is decided by the parameter which is most of all frequency-dependent. Very often, for example, the bandwidth will be limited by the impairment in the input match, that is, by the reduction in the TWR below a predetermined value. In other cases, the bandwidth may be limited by a change in the position of the maximum in the radiation pattern, beam broadening, an impairment in directivity, and so on. Cases have been ob-

served where the upper and lower limits of the bandwidth are controlled by different factors: for example, by the quality of match at the low-frequency end and by the distortion of the radiation pattern at the high-frequency end. A complete idea about the bandwidth of an antenna can only be formed after all design work has been completed and a prototype antenna has been tested.

By convention, the term 'narrowband' is assigned to antennas for which the bandwidth is less than 10% of the nominal frequency. Broadband antennas may have a bandwidth of 10 to 50%. Antennas with a bandwidth of up to several octaves are called superbroadband antennas. Lastly, if the ratio of the upper to the lower frequency limit is five-to-one and more, such an antenna may be classed as a frequency-independent one.

One more important parameter for transmitting antennas is the *maximum operating power*. It is usually limited by the electrical strength of the dielectrics used in the antenna and the electrical strength of the medium surrounding it (which is especially true of airborne antennas operating in the rarefied layers of the troposphere and the ionosphere). Also, in cases where a heavy power is fed to the input of the antenna, there is a risk of a thermal breakdown due to the heat build-up produced in the conductors and dielectrics by the transmitted r.f. power. Analysis of antennas for electrical and thermal strength constitutes a highly specific field of theory and lies outside the scope of this book.

Chapter Eight

The Receiving Antenna

8.1. The Equivalent Circuit of a Receiving Antenna

Consider two arbitrary antennas, *1* and *2*, located so that each is in the far-zone region of the other (Fig. 8.1). Let the normalized input impedances of the antennas in transmission be $\dot{z}_{A1}$ and $\dot{z}_{A2}$. Power transfer may in principle proceed in any one of two ways, namely: from the first to the second or from the second to the first. If antenna *1* is excited by a generator for which the normalized emf is $\dot{e}_1$ and which has a normalized internal impedance $\dot{z}_1$, the normalized input current will be

$$\dot{i}_1 = \dot{e}_1/(\dot{z}_{A1} + \dot{z}_1) \tag{8.1}$$

The radiation intensity of antenna *1* as measured at the location of antenna *2* (more accurately, at a point relative to which we know the directional phase characteristic of the second antenna in the

absence of the first) is, according to Eq. (7.25a), equal to

$$\dot{\mathcal{E}}_{21} = \dot{i}_1 (G_1 r_{A1}/4\pi)^{1/2} \dot{\mathbf{F}}_1 (\theta_1, \varphi_1) \exp(-j\beta R_0)/R_0 \qquad (8.2)$$

where G_1 = maximum gain of the first antenna
r_{A1} = real part of the input impedance
$\dot{\mathbf{F}}_1(\theta_1, \varphi_1)$ = sample of the complex vector radiation pattern of the first antenna in the direction of the second
R_0 = spacing between the points relative to which the phase characteristics of the antennas are specified

When the second antenna is placed in the field of a locally plane electromagnetic wave, Eq. (8.2), a normalized current $\dot{i}_{21}$, yet un-

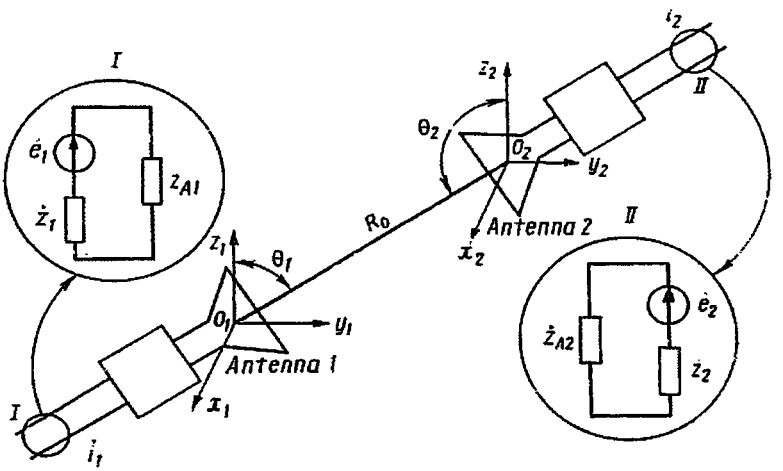


Fig. 8.1. Power transfer between two antennas

known, appears at the input plane of the second antenna loaded into a normalized impedance $\dot{z}_2$. On eliminating the current $\dot{i}_1$ between Eqs. (8.2) and (8.1), the emf can be written in an explicit form as

$$\dot{e}_1 = \frac{\dot{\mathcal{E}}_{21} (\dot{z}_{A1} + \dot{z}_1) R_0}{\dot{\mathbf{F}}_1 (\theta_1, \varphi_1) \exp(-j\beta R_0)} (4\pi/G_1 r_{A1})^{1/2} \qquad (8.3)$$

where the ratio of the vectors in the numerator and the denominator has a physical significance because these vectors are parallel.

In the other case, when a generator of normalized emf $\dot{e}_2$ and of normalized internal impedance $\dot{z}_2$ is connected to the second antenna and the first antenna loaded into $\dot{z}_1$ acts as a receiving one, we obtain

the emf in a similar way as

$$\dot{e}_2 = \frac{\dot{\vec{\mathcal{E}}}_{12}(\dot{z}_{A2}+\dot{z}_2)R_0}{\dot{F}_2(\theta_2, \varphi_2)\exp(-j\beta R_0)}(4\pi/G_2 r_{A2})^{1/2} \tag{8.4}$$

where G_2 = maximum gain of the second antenna

r_{A2} = real part of the input impedance $\dot{z}_{A2}$ in transmission

θ_2, φ_2 = angles of direction to the first antenna

The current $\dot{i}_{12}$ arising under the action of $\dot{e}_2$ at the input plane of the first antenna loaded into $\dot{z}_1$ is not yet known, either.

If the space between and around the two antennas is linear and isotropic, which means that it is free from biased, gyrotropic media, then, proceeding from the reciprocity theorem, we may write

$$\dot{e}_1/\dot{i}_{21} = \dot{e}_2/\dot{i}_{12}$$

This relation holds for power transfer in either direction. Invoking Eqs. (8.3) and (8.4), the above relation may be recast as

$$\frac{\dot{\vec{\mathcal{E}}}_{21}(\dot{z}_{A1}+\dot{z}_1)\dot{i}_{12}}{(G_1 r_{A1})^{1/2}\dot{F}_1(\theta_1, \varphi_1)} = \frac{\dot{\vec{\mathcal{E}}}_{12}(\dot{z}_{A2}+\dot{z}_2)\dot{i}_{21}}{(G_2 r_{A2})^{1/2}\dot{F}_2(\theta_2, \varphi_2)}$$

On collecting on each side of the equality the terms related only to one antenna, we obtain a constant

$$N = \frac{\dot{i}_{12}(\dot{z}_{A1}+\dot{z}_1)}{(G_1 r_{A1})^{1/2}\langle \dot{\vec{\mathcal{E}}}_{12}\dot{F}_1^*(\theta_1, \varphi_1)\rangle} = \frac{\dot{i}_{21}(\dot{z}_{A2}+\dot{z}_2)}{(G_2 r_{A2})^{1/2}\langle \dot{\vec{\mathcal{E}}}_{21}\dot{F}_2^*(\theta_2, \varphi_2)\rangle} \tag{8.5}$$

because arbitrary variations in the parameters of one antenna do not practically affect the parameters of the other antenna located in its far-zone region. Since the usual scalar product of two-dimensional vectors has the form

$$\langle \mathbf{ab}\rangle = \dot{a}_1\dot{b}_1^* + \dot{a}_2\dot{b}_2^*$$

it follows that the scalar products in (8.5) should be written out according to the rule

$$\langle \mathbf{ab}^*\rangle = \dot{a}_1\dot{b}_1 + \dot{a}_2\dot{b}_2$$

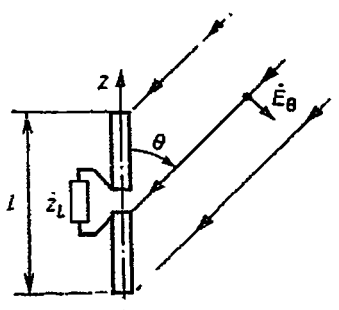


Fig. 8.2. Hertz dipole as a receiving antenna ($l \ll \lambda$)

In order to determine the constant N, let us consider the simplest antenna in the form of a Hertz dipole of length l and of load impedance Z_L (Fig. 8.2). Let the incident plane electromagnetic wave be linearly polarized with a vector

$$\dot{\mathbf{E}} = \mathbf{i}_0\dot{E}_{\theta 0}$$

where $\dot{E}_{\theta 0}$ is the electric field strength at the centre of the dipole. This wave produces across the terminals of the open-circuited dipole an emf

$$\mathscr{E}_{oc} = -\dot{E}_{\theta_0} l \sin \theta$$

which causes the following current to flow in the load:

$$\dot{I}_0 = \vec{\mathscr{E}}_{oc}/(\dot{Z}_L + \dot{Z}_{\Sigma 0}) = -(\dot{E}_{\theta_0} l \sin \theta)/(\dot{Z}_L + \dot{Z}_{\Sigma 0}) \qquad (8.6)$$

where

$$\dot{Z}_{\Sigma 0} = R_{\Sigma 0} + jX_{\Sigma 0}$$

is the input impedance of the dipole in transmission; the '—' sign indicates that this emf balances out the tangential component of the electric field strength on the dipole conductor. In order to normalize Eq. (8.6), the following should be done:

(1) Express sin θ in terms of the normalized radiation pattern of the dipole

$$\dot{F}(\theta) = j \sin \theta$$

(2) Using Eq. (7.21), change l to

$$h_l = (DR_{\Sigma 0}/Z_m \pi)^{1/2}$$

(3) Set the characteristic impedance of the feed line equal to Z_c and pass to the normalized impedances

$$\dot{Z}_L + \dot{Z}_{\Sigma 0} = (\dot{z}_L + \dot{z}_A) Z_c$$

Then Eq. (8.6) will take the form

$$I_0 (Z_c/2)^{1/2} = j\lambda (Dr_A)^{1/2} \dot{E}_{\theta_0} \dot{F}(\theta)/(2\pi Z_m)^{1/2} (\dot{z}_L + \dot{z}_A)$$

which is in effect identical to

$$\dot{i}_0 = \frac{j\lambda (Gr_A)^{1/2}}{\pi^{1/2} (\dot{z}_A + \dot{z}_L)} \langle \vec{\mathscr{E}} \dot{\mathbf{F}}^* \rangle \qquad (8.7)$$

where the directivity of the dipole has been replaced by its gain, $G = D\eta$, because it has been assumed that the dipole is free from ohmic losses and that the efficiency is $\eta = 1$.

On comparing Eqs. (8.7) and (8.5), we can see that

$$N = j\lambda/\pi^{1/2}$$

On substituting this expression for N in Eq. (8.5), we obtain the final expression for the normalized load current of a receiving antenna

$$\dot{i}_L = \frac{j\lambda (Gr_A)^{1/2}}{\pi^{1/2} (\dot{z}_A + \dot{z}_L)} \langle \vec{\mathscr{E}} \dot{\mathbf{F}}^* (\theta_0, \varphi_0) \rangle \qquad (8.8)$$

where $\dot{\vec{\mathscr{E}}}$ = radiation intensity vector of the plane electromagnetic wave incident on the antenna, having the dimensions of $(W/m)^{1/2}$

$\dot{F}(\theta, \varphi)$ = sample from the normalized complex vector radiation pattern of the antenna in the direction of wave arrival

$\dot{z}_A$ = normalized input impedance of the antenna

$\dot{z}_L$ = normalized load impedance

To sum up, any transmitting antenna may be used as a receiving one and treated as an *equivalent generator of normalized emf* $(W^{1/2})$

$$\dot{e}_{eq} = j\lambda (Gr_A/\pi)^{1/2} \langle \dot{\vec{\mathscr{E}}} \dot{\mathbf{F}}^* \rangle \tag{8.9}$$

and of *internal impedance equal to the input impedance of the antenna in transmission.*

The parameters needed to describe a receiving antenna include its complex vector radiation pattern function $\dot{\mathbf{F}}(\theta, \varphi)$, its gain $G = D\eta$, and its normalized input impedance $\dot{z}_A = r_A + jx_A$, with all of them definable in the transmission mode. Because the parameters $\dot{\mathbf{F}}(\theta, \varphi)$, G and $\dot{z}_A$ remain unchanged on switching from transmission to reception, there is no need to qualify which mode of operation is meant in our further discussion. The fact that one and the same antenna has the same parameters in transmission and reception was established by M.S. Neyman of the Soviet Union in 1935.

8.2. Polarization Relations in Reception

The scalar product $\langle \dot{\vec{\mathscr{E}}} \dot{\mathbf{F}}^* \rangle$ in Eqs. (8.8) and (8.9) is proportional to the scalar product of the polarization vector $\dot{\mathbf{p}}_e$ of the incident wave by the polarization vector $\dot{\mathbf{p}}_A$ of the receiving antenna

$$\langle \dot{\vec{\mathscr{E}}} \dot{\mathbf{F}}^* (\theta_0, \varphi_0) \rangle = \dot{\mathscr{E}} \dot{F} (\theta_0, \varphi_0) \dot{\xi} \tag{8.10}$$

where

$$\dot{\xi} = \langle \dot{\mathbf{p}}_e \dot{\mathbf{p}}_A^* \rangle, \quad 0 \leqslant |\dot{\xi}| \leqslant 1$$

may be termed the *field-strength polarization transfer function.*

In order to evaluate $\dot{\xi}$, let us express the vectors $\dot{\mathbf{p}}_e$ and $\dot{\mathbf{p}}_A$ on some common basis, say, in the coordinate system belonging to the receiving antenna, with unit vectors $\dot{\mathbf{i}}_\varphi$ and $\dot{\mathbf{i}}_\theta$. Suppose that the φ-component represents the principal polarization and the θ-com-

ponent, the cross-polarization. Then we may, on the basis of (7.14), write the following expressions for the polarization vectors:

$$\dot{\mathbf{p}}_A = \mathbf{i}_\varphi \alpha_A + \mathbf{i}_\theta (1 - \alpha_A^2)^{1/2} \exp(j\psi_A)$$

and

$$\dot{\mathbf{p}}_e = \mathbf{i}_\varphi \alpha_e + \mathbf{i}_\theta (1 - \alpha_e^2)^{1/2} \exp(j\psi_e)$$

On taking the sum of the products of the components of the vector $\dot{\mathbf{p}}_A$ by the respective components of the vector $\dot{\mathbf{p}}_e$, we get

$$\dot{\xi} = \alpha_A \alpha_e + [(1 - \alpha_A^2)(1 - \alpha_e^2)]^{1/2} \exp[j(\psi_A + \psi_e)] \qquad (8.11)$$

To begin with, let us establish the conditions under which there can be no reception. On equating to zero separately the real and imaginary parts of Eq. (8.11), we get

$$\psi_A + \psi_e = \pi$$

and

$$\alpha_A = (1 - \alpha_e^2)^{1/2}$$

This means that when $\dot{\xi}$ vanishes, the polarization vectors should have the following form:

$$\dot{\mathbf{p}}_A = \mathbf{i}\varphi \cos\tau + \mathbf{i}_\theta \sin\tau \exp(j\psi_A)$$

$$\dot{\mathbf{p}}_e = \mathbf{i}\varphi \sin\tau - \mathbf{i}_\theta \cos\tau \exp(-j\psi_A)$$

where

$$0 \leqslant \tau \leqslant \pi/2$$

and

$$-\pi/2 \leqslant \psi_A \leqslant \pi/2$$

are arbitrary parameters.

Simple analysis would show that such vectors $\dot{\mathbf{p}}_A$ and $\dot{\mathbf{p}}_e$ correspond to polarization ellipses which have the same axial ratios, have their major axes at right angles to each other, and rotate in the same direction. However, the same direction of rotation of the instantaneous values of the polarization vectors $\dot{\mathbf{p}}_A$ and $\dot{\mathbf{p}}_e$ on a common plane implies that for observers looking in the wake of the deproaching wave they appear rotating in the opposite directions. Hence we conclude that the polarization vectors belong to orthogonal polarizations. Thus, *a receiving antenna does not respond to the incident electromagnetic wave with a polarization orthogonal to its own polarization in the transmission mode.*

Let us define the conditions under which the quantity $\dot{\xi}$ has a maximum magnitude. For this to take place, it is essential, firstly, that

$$\psi_A + \psi_e = 0$$

which, on introducing the notation

$$\alpha_A = \cos \tau_1$$

and

$$\alpha_e = \cos \tau_2$$

gives

$$\dot{\xi} = \cos \tau_1 \cos \tau_2 + \sin \tau_1 \sin \tau_2 = \cos (\tau_1 - \tau_2)$$

Secondly, the magnitude of this expression has an absolute maximum equal to unity when $\tau_1 = \tau_2$, which implies that the polarization vectors should have the form

$$\dot{\mathbf{p}}_A = \mathbf{i}_\varphi \cos \tau + \mathbf{i}_\theta \sin \tau \exp (j\psi)$$

and

$$\dot{\mathbf{p}}_e = \mathbf{i}_\varphi \cos \tau + \mathbf{i}_\theta \sin \tau \exp (-j\psi)$$

where τ and ψ are arbitrary parameters. Such vectors $\dot{\mathbf{p}}_A$ and $\dot{\mathbf{p}}_e$ correspond to coincident polarization ellipses rotating in opposite directions. But the rotation of polarization ellipses in opposite directions on a common plane implies that for an observer looking in the wake of the deproaching wave the rotation appears to be in the same direction. Therefore we conclude that the polarization vectors belong to coincident elliptical polarizations. Thus, *a receiving antenna will collect all of the incident electromagnetic wave if its polarization is the same as the polarization of the antenna in the transmission mode.*

Equation (8.11) defining the field-strength polarization transfer function holds for any polarization basis specified by (7.15). What is essential is that in writing the vectors $\dot{\mathbf{p}}_A$ and $\dot{\mathbf{p}}_e$ use should be made of the same basis.

Polarization relations for circularly polarized antennas. Let there be a receiving antenna characterized by a right-hand (clockwise) circular polarization vector

$$\dot{\mathbf{p}}_A = 0.707 (\mathbf{i}_\varphi + j\mathbf{i}_\theta)$$

Furthermore, let an incident electromagnetic wave, described on the same basis, be characterized by an arbitrary polarization vector of the form

$$\dot{\mathbf{p}}_e = \mathbf{i}_\varphi \cos \tau + \mathbf{i}_\theta \sin \tau \exp (j\psi)$$

According to (8.11), the polarization transfer function is

$$\dot{\xi}_0 = 0.707 [\cos \tau + j \sin \tau \exp (j\psi)]$$

Let us turn the antenna about the direction of wave arrival through an angle β clockwise. In terms of the old $\mathbf{i}_\varphi$, $\mathbf{i}_\theta$ basis the polarization vector of the antenna after it has been turned takes the form

$$\dot{\mathbf{p}}_A(\beta) = 0.707\,[\mathbf{i}_\varphi(\cos\beta + j\sin\beta) + \mathbf{i}_\theta(j\cos\beta - \sin\beta)$$
$$= \dot{\mathbf{p}}_A \exp(j\beta)$$

As the receiving antenna is turned through some angle, its polarization transfer function takes the form

$$\dot{\xi}(\beta) = \langle \dot{\mathbf{p}}_e \dot{\mathbf{p}}_A(\beta) \rangle = \dot{\xi}_0 \exp(j\beta)$$

That is, rotation of a circularly polarized antenna about the direction of the communication link through an angle β does not change the received voltage in magnitude and only brings on an excess phase shift numerically equal to the rotation of the antenna.

Circularly polarized antennas are widely used for communication with aircraft: With coincident circular polarizations, the magnitude of the received voltage does not depend on the rotation of the antenna about the direction of the communication link, and the polarization power transfer function is $|\dot{\xi}|^2 = 1$. The fact that the rotation of circularly polarized radiators is equivalent to introducing an excess phase shift in the communication channel is utilized in mechanical microwave polarization phase shifters. It can also be utilized in experimental studies of circularly polarized phased antenna arrays to model the phase shifts of the excitation currents in the radiators of an array.

As regards radar, it is important that when *a common circularly polarized antenna is used for both transmission and reception, it provides for polarization-related cancellation of echo signals from objects having a symmetrical shape*, such as spherical rain drops. When circularly polarized waves are reflected from objects having a symmetrical shape, including an infinite metal plane, the direction in which the polarization vector rotates relative to an observer looking in the wake of the deproaching wave is changed. As a result, the reflected wave is orthogonally polarized with respect to the antenna, and the latter does not respond to the wave. If, on the other hand, echos come from unsymmetrically shaped objects, such as extended conductors, the reflected wave contains two circularly polarized components rotating in opposite directions — one of these components is collected by the antenna and produces a useful signal.

8.3. Load Power of a Receiving Antenna. Effective Area

If we use Eqs. (8.8) and (8.10), we may write the useful signal power developed in the load of a receiving antenna as

$$P_s = |\dot{i}_L|^2 r_L = |\vec{\dot{\mathscr{E}}}|^2 (\lambda^2 G/4\pi) F^2(\theta_0, \varphi_0) |\dot{\xi}|^2 \frac{4 r_L r_A}{|\dot{z}_A + \dot{z}_L|^2}$$

where $|\vec{\dot{\mathscr{E}}}|^2$ = magnitude of the Poynting vector of the wave incident on the antenna
G = maximum antenna gain
$F^2(\theta_0, \varphi_0)$ = sample from the normalized power radiation pattern in the direction of wave arrival
$|\dot{\xi}|^2$ = polarization power transfer function
$\dot{z}_A = r_A + jx_A$ = normalized input impedance of the antenna
$\dot{z}_L = r_L + jx_L$ = normalized load impedance referred to the antenna input

The last factor in the above equation takes care of the mismatch between the antenna and the load. If the antenna is properly matched to its feed line, and the reflection coefficient of the load is $\dot{\rho}_L$, then

$$4 r_L r_A / |\dot{z}_A + \dot{z}_L|^2 = 1 - |\dot{\rho}_L|^2 \text{ at } \dot{z}_A = 1$$

and the equation for the received power takes the form

$$P_s = |\vec{\dot{\mathscr{E}}}|^2 S_{eff} F^2(\theta_0, \varphi_0) \eta |\dot{\xi}|^2 (1 - |\dot{\rho}_L|^2) \tag{8.12}$$

where

$$S_{eff} = \lambda^2 D/4\pi$$

has the dimensions of area and is called the *effective receiving cross-section* or *area* of an antenna. The concept of the effective area is closely related to the maximum power that a receiving antenna can collect from the incident electromagnetic wave. As follows from the structure of Eq. (8.12), four conditions can be identified that are essential for the load power of an antenna to be a maximum. These conditions are as follows.

(1) The maximum of the radiation pattern should be aligned with the direction of wave arrival, such that $F^2(\theta_0, \varphi_0) = 1$.

(2) Ohmic losses in the antenna and in the matching network should be kept to a minimum, so that in the limit $\eta = 1$.

(3) The polarization of the antenna should be precisely the same as that of the incident wave, such that $|\dot{\xi}|^2 = 1$.

(4) The antenna should be properly matched to the load, so that $\dot{\rho}_L = 0$ (more generally, this condition may be written as $\dot{z}_L = \dot{z}_A^*$).

When all of the four conditions are satisfied, the maximum power received by an antenna is equal to the product of the Poynting vector $|\dot{\vec{\mathcal{E}}}|^2$ of the incident wave by the effective area of the antenna, S_{eff}. Thus, the effective area of an antenna can be interpreted as the area of the front of a plane electromagnetic wave from which the antenna collects and transfers to a matched load the received power, provided that the maximum of the radiation pattern is aligned with the direction of wave arrival, that there are no ohmic losses in the antenna, and that the incident wave and the antenna have the same polarization.

The relation connecting the directivity and the effective area of an antenna thus found

$$D = 4\pi S_{eff}/\lambda^2 \tag{8.13}$$

is among the fundamental ones in antenna theory. As will be shown later, for aperture-type antennas the effective area is directly related to the size of the radiating aperture and does not usually exceed its surface area. For dipole and slot antennas (including elementary dipoles), the effective area cannot be expressed directly in terms of radiator dimensions and is introduced in a purely formal manner.

8.4. The Noise Temperature of a Receiving Antenna

In order to be able to say whether a weak signal can be picked up by a given receiver, it is essential to compare the power carried by such a signal with the total noise power present at the receiver input. Some of the noise is associated with the antenna, whereas the remainder has its origin in the receiver. The noise produced by an antenna may be classed into *external* and *internal*. The prevalent contribution comes from the external noise picked up by the antenna from the surroundings; it owes its origin to man-made interference, lightning discharges, emissions from galactic sources and thermal radiation from the Earth's surface, the troposphere, and the ionosphere. A smaller contribution comes, as a rule, from internal noise which has its origin in the thermal motion of electrons in the non-ideal conductors and dielectrics used in the antenna and the feeder system.

Since external noise and internal noise are quite equivalent in spectral composition and interfering action, it is customary to specify their aggregate effect in terms of a single parameter called the *noise temperature of an antenna*. It is given the symbol of T_A and is measured in kelvins. The noise temperature is assigned to the internal resistance of the antenna and offers an opportunity to find

the antenna noise power P_{nA} per unit bandwidth Δf applied to a matched receiver by what is known as *Nyquist's formula* (or *theorem*)

$$P_{nA} = kT_A\Delta f$$

where k is Boltzmann's constant equal to 1.38×10^{-23} W Hz^{-1} K^{-1}.

In this way, all forms of noise (both picked up and produced by an antenna) are equivalently replaced by the thermal noise in its input resistance at a hypothetical temperature, T_A. This is quite analogous to what is done in the case of receivers where all forms of internal receiver noise are replaced by an equivalent noise associated with the input resistance to which is assigned the equivalent noise temperature of the receiver

$$T_r = T_0 (N_r - 1)$$

where T_0 = standard ambient temperature under terrestrial conditions, equal to 288K

N_r = receiver noise factor

The difference consists in that in the case of a receiver all forms of noise are equivalently referred to the input, whereas in the case of an antenna they are referred to the output. This simplifies the calculation of the power contributed by the signal and noise when an antenna is operating jointly with a receiver. The total noise power of a radio system at the receiver input is then equal to

$$P_n = P_{nA} + P_{nr} = k\,\Delta f\,(T_A + T_r)$$

The signal power available at the input of a receiver matched to an antenna is expressed as the product of the signal Poynting vector, Π_s, by the effective area and efficiency of the antenna. Under conditions of a perfect match in terms of polarization and impedance, the signal-to-noise ratio (m^2/K) is given by

$$S/N = P_s/(P_{nA} + P_{nr}) = (\Pi s/k\Delta f)\,[S_{eff}\eta/(T_A + T_r)]$$

Sometimes, the ratio $S_{eff}\eta/(T_A + T_r)$ is referred to as the *sensitivity of a receiving antenna system* (m^2/K).

The total noise temperature of an antenna is the sum of two components: the contribution by thermal fluctuation noise T_{Af} and the noise picked up from external sources, $T_{A\Sigma}$:

$$T_A = T_{Af} + T_{A\Sigma}$$

The contribution to the noise temperature of an antenna due to thermal fluctuations in non-ideal conductors and dielectrics is found as follows. According to Nyquist's theorem, the normalized noise emf appearing at the output of an antenna due to thermal fluctuations is

$$\overline{e_f} = (4kT_0\Delta f r_{loss})^{1/2}$$

where

$$r_{loss} = r_A - r_\Sigma$$

is the resistive component of the input impedance of the antenna, taking care of the ohmic losses, and T_0 is 288K. The noise power at the input of a matched receiver (with an input impedance equal to $\dot{z}_{in} = r_A - jx_A$) corresponding to the above emf is

$$\begin{aligned} P_{nf} &= |\bar{e}_f|^2 r_A/(r_A + r_\Sigma)^2 \\ &= k\Delta f T_0 [(r_A - r_\Sigma)/r_A] \\ &= k\Delta f T_0 (1 - \eta) \end{aligned}$$

where

$$\eta = r_\Sigma/r_A$$

is the efficiency of the antenna. Hence, we obtain a simple equation to describe the contribution to the noise temperature of an antenna by ohmic losses:

$$T_{Af} = T_0 (1 - \eta) \tag{8.14}$$

Now let us turn to the contribution due to the reception of external noise, $T_{A\Sigma}$. Suppose that this noise is equivalent to the thermal radiation of an absolutely black surface enclosing the antenna, say, a sphere of a radius exceeding the distance to the far-zone region of the antenna. If the surface of the black sphere is uniformly heated to a brightness temperature T_b, the flux density of the radiated electromagnetic power at a wavelength λ over the bandwidth Δf per unit solid angle will, according to the Rayleigh-Jeans radiation formula, be equal to

$$d\Pi_n/d\Omega = 2kT_b\Delta f/\lambda^2$$

It is legitimate to invoke the Rayleigh-Jeans radiation formula because the energy of a quantum of electromagnetic energy (that is, of a photon) at radio frequencies is substantially smaller than kT.

The noise power dissipated in the load of a matched receiving antenna located at the centre of the enclosing sphere is

$$P_{n\Sigma} = 0.5 \oint\limits_{\Omega=4\pi} (d\Pi_n/d\Omega)\, S_{eff}\eta F^2(\theta, \varphi)\, d\Omega$$

where the factor 0.5 indicates that the reception of a randomly polarized field involves the loss of a half of the power flux density with a polarization orthogonal to the own polarization of the antenna. Recalling that

$$S_{eff} = \lambda^2 D/4\pi$$

and invoking Eq. (7.18) which gives the directivity of an antenna we obtain

$$P_{n\Sigma} = kT_b \Delta f \eta$$

and

$$T_{n\Sigma} = \eta T_b$$

Thus, when the brightness temperature is uniformly distributed over a large sphere, the noise temperature of the antenna is independent of the shape of its radiation pattern and is solely decided by its brightness temperature and efficiency.

If, on the other hand, the surface of the enclosing sphere is heated nonuniformly and the brightness temperature is distributed as $T_b(\theta, \varphi)$, the contribution to the noise temperature due to the reception of thermal radio emission is given by

$$T_{A\Sigma} = (D\eta/4\pi) \oint\limits_{\Omega=4\pi} T_b(\theta, \varphi) F^2(\theta, \varphi) \, d\Omega \qquad (8.15)$$

where $D\eta$ = antenna gain
$F^2(\theta, \varphi)$ = power radiation pattern
$T_b(\theta, \varphi)$ = angular distribution of the brightness temperature over the sphere enclosing the antenna

Now the noise temperature of the antenna strongly depends on its radiation pattern and its orientation in space.

Suppose that the radiation pattern of an antenna has a very narrow main lobe and negligibly small side lobes such that, in the limit,

$$\lim [F^2(\theta, \varphi)] = 4\pi\delta(\theta - \theta_0, \varphi - \varphi_0)/D$$

where $\delta(x, y)$ = two-dimensional Dirac delta-function
θ_0, φ_0 = direction of the main maximum

Then, according to Eq. (8.15), the noise temperature of the antenna is

$$T_{A\Sigma} = \eta \oint\limits_{\Omega=4\pi} T_b(\theta, \varphi) \delta(\theta - \theta_0, \varphi - \varphi_0) \, d\Omega = \eta T_b(\theta_0, \varphi_0)$$

Thus, the angular distribution of brightness temperature can be measured by consecutively scanning the surrounding space by a narrowbeam antenna. This form of study is carried out on a regular basis in the various frequency bands and constitutes the subject matter of a special science called radio astronomy.

On the other hand, analysis of Eq. (8.15) suggests the possibility of using galactic 'point-like' (discrete) sources of noise radio emission for the purpose of measuring the radiation pattern of ground-based narrowbeam antennas. In this case, the distribution $T_b(\theta, \varphi)$ in the vicinity of a discrete source may be approximated by the Dirac delta-function, and the record of noise power at the output of a receiver (a radiometer), obtained as the discrete source passes through

the direction of maximum radiation, will duplicate the radiation pattern of the antenna. This form of antenna radiation pattern measurements with the aid of extraterrestrial radio emission sources is quite feasible in the UHF and SHF bands where the troposphere and the ionosphere do not markedly absorb radio waves.

Generally, the effect of radio wave absorption on the noise temperature of an antenna is accounted for as follows. The radio brightness of extraterrestrial sources, $T_\infty(\theta, \varphi)$, is replaced by the quantity

$$T_b(\theta, \varphi) = T_\infty(\theta, \varphi)\,\eta(\theta, \varphi) + T_{atm}[1 - \eta(\theta, \varphi)]$$

where $\eta(\theta, \varphi)$ = total coefficient of power transmission of plane radio waves through all the absorbing layers of the atmosphere

T_{atm} = averaged physical temperature of the absorbing layers in kelvins

The first term shows that the absorption brings down the noise power from sources located beyond the absorbing layers. The second term represents the own fluctuation noise of the absorbing layers and is analogous in structure to Eq. (8.14).

Noise sources. In the LF, MF and HF bands, the predominant role is played by impulse noise produced by lightning and industrial discharges. The current spectral density of short electric pulses is inversely proportional to frequency. Therefore, the respective noise power and noise temperature decrease in inverse proportion to the square of frequency. The power spectral density of noise due to distant thunderstorms at the point of reception shows a more involved frequency dependence in view of the likely reflections from the ionosphere. Moreover, the intensity of interference due to distant thunderstorms is subject to daily and seasonal variations. The level of atmospheric noise and interference in a particular locality is shown on special maps compiled for the four seasons of the year and six intervals spanning between them 24 hours. Such maps show lines of constant noise temperature for a short vertical antenna at a standard frequency of 1 MHz. Conversion to the operating frequency of a real radio system is facilitated by additional frequency charts. In the middle region of the Soviet Union, the noise temperature due to atmospherics, as measured at a frequency of 1 MHz, is 60 dB relative to 288K, which works out to 3×10^8K. At a frequency of 30 MHz, the figure falls to 300K and tends to zero as the frequency is raised still more. To a first approximation, it may be deemed that sources of thunderstorm noise are distributed in space uniformly, which implies that the noise temperature is independent of the shape of the radiation pattern of an antenna.

Because the external noise temperature is high at low and medium frequencies, the effect of antenna efficiency on receiver sensiti-

vity may prove negligibly small. To demonstrate, the signal-to-noise ratio takes the form

$$S/N = \frac{\Pi_s \lambda^2}{4\pi k \Delta f} \left[\frac{\eta D}{T_0 (1-\eta) + \eta T_b + T_r} \right]$$

where ηD = antenna gain
T_b = brightness temperature of atmospherics, as found from maps
T_r = receiver noise temperature

If the inequality

$$\eta T_b \gg T_0 (1 - \eta) + T_r$$

is satisfied, the signal-to-noise ratio is proportional to D/T_b and is independent of the antenna efficiency. Hence it is possible to maintain consistent reception at medium and low frequencies, using small-sized loop and wire antennas with an efficiency of the order of 10^{-4} or even less. The low signal voltage at the output of such an 'ineffectual' antenna can be made up for by increasing the receiver gain (sensitivity).

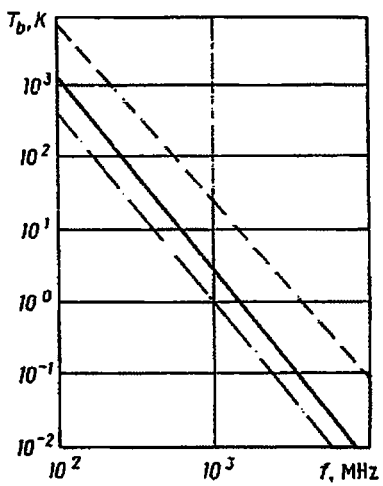


Fig. 8.3. Brightness temperature of the sky at different frequencies

At frequencies upwards of 30 MHz, the level of atmospherics falls abruptly because the noise radio emission produced by thunderstorm sources cannot propagate by multiple reflections from the ionosphere. On the other hand, galactic noise begins to be felt because the ionosphere is 'transparent' to it at these frequencies. Radio astronomers have compiled detailed maps showing the brightness temperature of galactic sources which are at their strongest near the Galactic centre and the Galactic poles. The manner in which the intensity of galactic radiation depends on frequency is customarity characterized in terms of the averaged noise temperature of an isotropic antenna in free space at various frequencies. This temperature is computed on the basis of radio brightness maps of the sky (the full line in Fig. 8.3). The dashed and dot-dashed lines in Fig. 8.3 represent the average brightness temperature of the 'hot' and 'cold' regions of the sky, respectively.

When ground-based antennas are compared in terms of their noise temperature, it is customary to orient them so that the main lobe of the radiation pattern points at a 'cold' region of the sky, usually

at zenith. In the circumstances, the largest contribution to the noise temperature is due to the fact that the backward and side lobes of the antenna pick up thermal radio emission from the Earth and emissions from galactic sources. Therefore, in order to minimize the noise temperature of narrow-beam antennas operating in the UHF and SHF bands, it is essential to reduce their backward and side lobes. With a properly designed antenna, the contribution to the noise temperature coming from the reception of radio emission radiated by the Earth and galactic sources can be reduced to anywhere between 20 and 5K. Against this low background, a prominent role is played by fluctuation noise, $T_0(1 - \eta)$. For example, at $\eta = 0.95$ this contribution is around 15K and turns out to be comparable with $T_{A\Sigma}$. Thus, the efficiency of low-noise UHF and SHF receiving antennas must be as high as practicable.

8.5. Mutual Impedance of Spatially Separated Antennas

Let us go back to power transfer between two antennas (see Fig. 8.1). Let also agree to treat the inputs of the antennas as the ports of some equivalent reciprocal two-port which includes the two antennas and the space between them. As before, we take it that each antenna is located in the far-zone field of the other. If the antenna inputs are matched to the feed lines ($\dot{z}_{A1} = \dot{z}_{A2} = 1$) and the presence of the other antenna in an open-circuit condition does not disturb this match, the equivalent two-port may be described by the following normalized impedance matrix:

$$\begin{bmatrix} \dot{v}_1 \\ \dot{v}_2 \end{bmatrix} = \mathbf{Z} \begin{bmatrix} \dot{i}_1 \\ \dot{i}_2 \end{bmatrix}$$
$$\mathbf{Z} = \begin{bmatrix} 1 & \dot{z}_{12} \\ \dot{z}_{12} & 1 \end{bmatrix} \tag{8.16}$$

where the off-diagonal elements are the *mutual impedances* of the antennas.

Let a generator of normalized emf $\dot{e}_1$ and of normalized internal impedance $\dot{z}_1$ be connected to the first antenna, and the second antenna be a receiving one with a load $\dot{z}_2$. The set of Kirchhoff loop equations corresponding to (8.16) has the form

$$\begin{aligned} \dot{v}_1 &= \dot{i}_1 + \dot{i}_2 \dot{z}_{12} \\ 0 &= \dot{i}_1 \dot{z}_{12} + \dot{i}_2 (1 + \dot{z}_2) \end{aligned} \tag{8.17}$$

where

$$\dot{v}_1 = \dot{e}_1 - \dot{i}_1 \dot{z}_1$$

is the normalized voltage at the input of the first antenna. As follows from the second line in (8.17), the normalized mutual impedance may be written as

$$\dot{z}_{12} = -\dot{i}_2 (1 + \dot{z}_2)/\dot{i}_1$$

On writing $\dot{i}_2$ in accordance with Eq. (8.8) and using Eq. (8.2), we get

$$\dot{z}_{12} = -j (G_1 G_2)^{1/2} \langle \dot{\mathbf{F}}_1 \dot{\mathbf{F}}_2^* \rangle \exp(-j\beta R)/(\beta R) \qquad (8.18)$$

where G_1 = maximum gain of the first antenna
G_2 = maximum gain of the second antenna
$\langle \dot{\mathbf{F}}_1 \dot{\mathbf{F}}_2^* \rangle$ = scalar product of the samples from the normalized complex vector radiation patterns, taken in the direction joining the centres of the antennas

For the first time, Eq. (8.18) was derived by G.T. Markov of the Soviet Union in 1948.

The physical significance of Eq. (8.18) is clear enough. With two antennas located in each other's far-zone field, the magnitude of their mutual impedance is a slowly decreasing function of the form $1/\beta R$ (provided the radiation pattern and the polarization transfer function are non-zero). The phase of the mutual impedance varies in a manner typical of an electromagnetic wave travelling in free space. Equation (8.18) is an *asymptotic* one; this means that it exactly holds when the spacing between the antennas tends to infinity. In practice, it may be used when the spacing between antennas exceeds the size of the transition zone of each antenna, that is, when

$$R \geqslant 2L^2/\lambda$$

where L is the size of the larger antenna.

Proceeding from the mutual impedance of two antennas, it is possible to determine the response of one to a change in the load of the other. As follows from the definition of the Z matrix, Eq. (8.16), each antenna is ideally matched when the other is open-circuited at the output. As follows from Eq. (8.17), connection of a load impedance $\dot{z}_2$ to the output of the receiving antenna causes the input impedance of the transmitting antenna to change as follows:

$$\dot{z}_{in} = \dot{v}_1/\dot{i}_1 = 1 - \dot{z}_{12}^2/(1 + \dot{z}_2)$$

and to become practically mismatched at $|\dot{z}_{12}| \leqslant 0.1$.

8.6. Power Transfer Between Two Antennas

Equations (7.25) for the radiation intensity of an arbitrary antenna and Equations (8.12) for the power dissipated in the load of a receiving antenna can be combined and used to calculate the 'overall' power transfer function between two antennas each of which is located in the other's far-zone field in free space.

On recalling that the directivity and effective area of an antenna are connected as

$$S_{eff} = D\lambda^2/4\pi$$

we may write the power carried by the valid signal at the output of the receiving antenna, P_s, in three alternative forms as follows:

$$P_s/P_{xmtr} = \frac{\eta_1\eta_2 F_1^2 F_2^2 \mid \dot{\xi} \mid^2 (1 - \mid \dot{\rho}_1 \mid^2)(1 - \mid \dot{\rho}_2 \mid^2)}{R^2} \begin{cases} D_1 S_{eff2}/4\pi & (8.19a) \\ D_1 D_2 \lambda^2/16\pi^2 & (8.19b) \\ S_{eff1} S_{eff2}/\lambda^2 & (8.19c) \end{cases}$$

where P_{xmtr} = transmitter incident-wave power at the input of the transmitting antenna
R = spacing between the antennas
D_i = directivity of one of the antennas
η_i = antenna efficiency
$S_{eff\,i}$ = effective area
F_i^2 = sample from the power radiation pattern in the direction of the other antenna
$\dot{\rho}_i$ = reflection coefficient
$\mid \dot{\xi} \mid^2$ = polarization power transfer function
i = ith antenna (it is immaterial which antenna is transmitting or receiving)

Equations (8.19) refer to *ideal radio transmission* because they ignore the effect of the medium surrounding the antenna (the Earth, the atmosphere, and obstacles along the path of wave propagation). Most often, these equations are used in the design of radio links between ground-based stations and aircraft or space vehicles within the line-of-sight distance.

As an example, consider an illustrative case. Let the parameters of a ground-based antenna be fixed in advance and assume that the beam control system keeps the maximum of its radiation pattern always pointing at a flying craft. Then the power transfer factor between the two antennas turns out to be proportional to the product

$$D_a F_a^2(\theta, \varphi) \mid \dot{\xi}(\theta, \varphi, \beta) \mid^2$$

where D_a = maximum directivity of the on-board antenna
$F_a^2(\theta, \varphi)$ = sample from the normalized power radiation pattern of the on-board antenna

$|\dot{\xi}(\theta, \varphi, \beta)|^2$ = polarization power transfer function which is a function of the angles θ and φ that define the direction to the ground-based station and the angle β defining the angle that the on-board antenna makes with the ray (the direction of the communication link)

As the craft manoeuvres, all the three angles change, and so does the distance between the flying craft and the ground station. Two distinct situations may arise in the circumstances.

(1) The flight path of the craft and its attitude (orientation in space) are known in advance (as when placing an artificial earth satellite or a space probe in orbit).

(2) The flying craft follows a varying path and changes its attitude in an unpredictable manner so that all the three angles may change in a random manner between the maximum limits.

In case (1), one knows the position of the ray (the direction of the communication link) in the craft's local coordinates and the distance R at each instant of time, and the solid angles encompassing all the likely directions of the link take up only some part of the complete solid angle. This offers an opportunity to formulate the requirements for the shape of the radiation pattern and the polarization of the on-board antenna which, when satisfied, will assure the desired transfer function at any instant. Following that, the designer can readily choose a particular configuration for the on-board antenna that will meet these requirements wholly or in part. An essential point here is that the requirements to be met by the vector radiation pattern involve only a part of the complete solid angle, so that they can be met in principle and the designer is in a position to specify the realizable directivity of the antenna in accordance with Eq. (7.18).

In case (2), an entirely different situation exists. Here, for an uninterrupted communication one will need an isotropic antenna which has a constant state of polarization. However, antenna theory has proved the following theorem: *If the field-strength radiation pattern has no nulls, the polarization of radiation depends in a substantial manner on direction, and the complete solid angle, 4π, will always contain at least one direction along which the axial ratio will take on any predetermined value such that* $-1 \leqslant r \leqslant 1$.

In the case of radio communication with an unoriented craft, a corollary of the above theorem states that there is at least one combination of the three angles θ, φ, β for which power transfer between the antennas reduces to zero, and the communication is interrupted. The following question arises here: How can one, in a situation like this, say how much a given real antenna differs from an isotropic radiator? One approach is to use the distribution function of the

random variable $D_a F_a^2(\theta, \varphi) \mid \dot{\xi}(\theta, \varphi, \beta) \mid^2$ in a three-dimensional space bounded by the angles θ, φ, and β so as to compute the total probability that for a great number of independent combinations of the angles θ_i, φ_i, and β_i the condition

$$D_a F_a^2(\theta_i, \varphi_i) \mid \dot{\xi}(\theta_i, \varphi_i, \beta_i \mid^2 \geqslant D_{eff}$$

(where D_{eff} is some fixed level) will be satisfied.

Quite often, this statistical funcion is referred to as the *communication probability* and denoted as $P\{D_{eff}\}$. This implies the application of the ergodicity principle by which the integral distribution $P\{D_{eff}\}$ obtained in independent trials will not change if the change in the combination of the angles θ, φ, β caused by the craft's manoeuvres in time is a stationary, random process in which any combination of angles has an equal probability of occurrence. When the transmitter power and the spacing between the antennas are fixed in advance and D_{eff} is construed as the least allowable value at which it is still possible to detect the valid signal in background noise, the probability $P\{D_{eff}\}$ does represent the probability of maintaining communication.

The value of $P\{D_{eff}\}$ can be found on a computer, using the *Monte Carlo method*. Using the known vector radiation pattern of the on-board antenna and the specified polarization vector of the ground-based antenna, one compiles a sub-program for the computation of the function

$$D_{eff,i} = D_a F_a^2(\theta_i, \varphi_i) \mid \dot{\xi}(\theta_i, \varphi_i, \beta_i) \mid^2$$

Then it is complemented by a sub-program to generate triads of random numbers θ_i, φ_i, β_i (possibly along with the required joint probability density function for orientation in space, or craft attitude). These sub-programs consecutively compute a number M of values of the function $D_{eff,i}$ and determine the number of cases, $K(D_{eff})$, in which $D_{eff,i}$ is greater than or equal to D_{eff}, where D_{eff} is a fixed level. For each value of D_{eff}, the ratio $K(D_{eff})/M$ is an estimate of the communication probability, $P\{D_{eff}\}$. The computational accuracy improves with increasing M. For example, for $M = 100$, the found value of $P\{D_{eff}\}$ lies in the $\pm 10\%$ interval with a probability of 90%. The Monte Carlo method may also be used to determine $P\{D_{eff}\}$ by experiment, without first measuring the radiation pattern of the on-board antenna.

Three typical examples illustrating the calculation of the communication probability for aircraft antennas with the simplest toroidal radiation pattern are shown in Fig. 8.4. Curve *1* represents the communication probability when the aircraft antenna is a Hertz dipole

and the ground-based station uses a linearly polarized antenna. Naturally, the maximum value of D_{eff} in this case is 1.5 because the polarization power transfer function turns to unity when the dipole lies in the E plane of the ground antenna. However, for a communication probability of 80%, the effective directivity of a Hertz dipole is a mere 0.06. The situation can be improved by using, on the ground, a circularly polarized antenna (curve *2*). Here, for a communication probability of 80%, the effective directivity of the Hertz

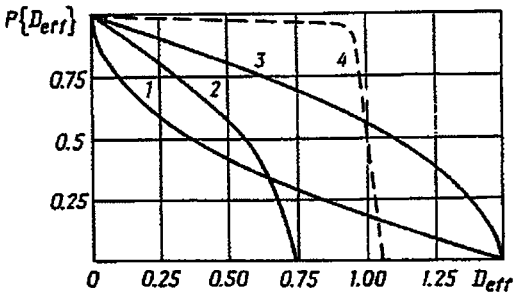


Fig. 8.4. Computed communication probability curves

dipole will be as high as 0.27. Still, the maximum value of D_{eff} cannot exceed 0.75, because the polarization power transfer function is only 0.5.

For a better power transfer, the aircraft should likewise use a circularly polarized antenna. For such an antenna with a radiation pattern of the form of $F^2(\theta, \varphi) = \sin^2\theta$, the communication probability is represented in Fig. 8.4 by curve *3*. With a communication probability of 80%, the effective directivity is now 0.54, or nine times the figure for curve *1* (both antennas are linearly polarized). For comparison, Fig. 8.4 also shows an idealized communication probability curve, *4*, which is the goal that the designer should strive to achieve when designing an optimal airborne antenna. Unfortunately, the unavoidable absence of communication along at least one direction does not permit this curve to be ideally rectangular.

If, however, there is a need for a 100% communication probability in a radio system, one has either to pin-point the radiation peak of the airborne antenna at the desired ground-based station or to receive separately orthogonally polarized signals on the ground and to employ an airborne antenna with a field-strength radiation pattern which has no radiation nulls.

Chapter Nine

Dipole and Slot Antennas

9.1. The Electric Dipole. Current and Charge Distributions

An *electric dipole* is an electromagnetic wave radiator in the form of two thin conductors of lengths l_1 and l_2 and of radius a, arranged as shown in Fig. 9.1*a* and driven by an r.f. generator at their junction point. Dipoles are widely used as stand-alone antennas or as elements of more elaborate antenna systems.

The generator emf gives rise in the dipole to longitudinal currents which are distributed on its surface in such a way that the electromagnetic field they set up satisfies Maxwell's equations, the boundary conditions at the conductor surface, and the condition for radiation at infinity. Owing to axial symmetry the currents on the side surface of the dipole have only longitudinal components, $\dot{J}^e_z$, whereas the currents at the ends only have radial components, $\dot{J}^e_\rho$. At the cylindrical surface of radius a, which covers the gap between the dipole halves, fictitious longitudinal electric currents co-exist with fictitious azimuthal magnetic surface currents, $\dot{J}^m_\varphi$, which equivalently replace the internal gap region along with the driving generator (see the principle of equivalence of field sources in the Appendix).

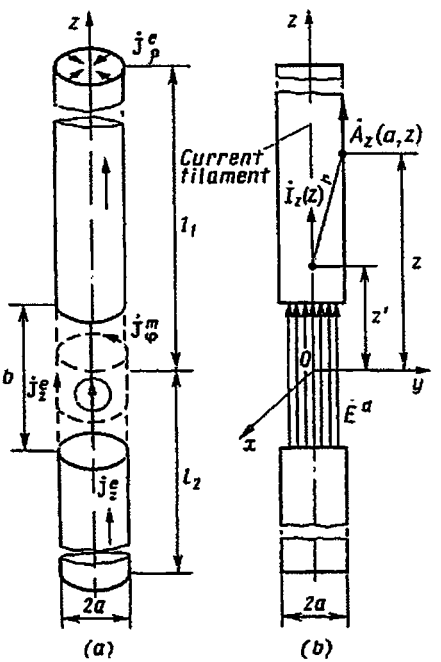


Fig. 9.1. Electric dipole (*a*) and its electrodynamic model (*b*)

In a theoretical study of the dipole it is above all essential to establish the distribution of radiating currents on its surface, that is, to solve the *internal problem*. Following that, one may proceed to tackle the *external problem*, that is, to determine the radiation pattern and other parameters of the dipole.

Integral equations for the current distribution on a dipole. For thin

dipoles ($a \ll l_1, l_2, \lambda$), an electrodynamic model for the internal problem is built on the basis of the following assumptions.

1. The surface electric currents $\dot{J}_z^e$ along with the magnetic equivalent currents $\dot{J}_\varphi^m$ in the gap region are replaced by an infinite filament of longitudinal electric current,

$$\dot{I}_z(z) = 2\pi a \dot{J}_z^e(z)$$

lying at the axis of the dipole. This current is deemed to be continuous in the gap region and to vanish at the dipole ends. The currents at the end faces are ignored.

2. The tangential component of the electric field vector, $\dot{E}_z(z)$, due to the current filament on the side surface of an imaginary cylinder enclosing the current filament at $\rho = a$ vanishes everywhere except the gap region of width b, where it is equated to some driving function $\dot{E}^d(z)$. For narrow gaps, the function $\dot{E}^d(z)$ may be deemed constant.

In accordance with (7.1), the unknown current distribution $\dot{I}_z(z)$ gives rise at the imaginary side surface of the dipole to a vector potential (see Fig. 9.1*b*) which has only one component

$$\dot{A}_z^e(z) = \frac{1}{4\pi} \int_{z'=-l_2}^{l_1} \dot{I}_z(z')\, \dot{\mathscr{K}}(z-z')\, dz' \tag{9.1}$$

where

$$\dot{\mathscr{K}}(z-z') = \exp(-j\beta r)/r$$

is the *kernel* of the integral, which is a function of the difference in coordinate z between the points of integration and observation because

$$r = [(z-z')^2 + a^2]^{1/2}$$

Using Eq. (A.4) for conversion from the electric-current vector potential to the electric field vector and imposing the boundary condition on the behaviour of $\dot{E}_z$ on the side surface of the dipole ($\rho = a$), we obtain

$$E_z|_{\rho=a} = \frac{1}{j4\pi\omega\varepsilon_a}\left(\beta^2 + \frac{d^2}{dz^2}\right) \int_{z'=-l_2}^{l_1} \dot{I}_z(z')\, \dot{\mathscr{K}}(z-z')\, dz'$$

$$= \begin{cases} 0 & \text{for } |z| > b/2 \\ \dot{E}^d & \text{for } |z| < b/2 \end{cases} \tag{9.2}$$

The unknown function $\dot{I}_z(z)$ in Eq. (9.2) is operated upon by the integration operator with the kernel $\dot{\mathscr{K}}(z - z')$, and this operator is in turn acted upon by the differentiation operator $(\beta^2 + d^2/dz^2)$. Therefore, Eq. (9.2) is an *integrodifferential equation* in $\dot{I}_z(z)$ with the kernel $\dot{\mathscr{K}}(z - z')$. For the first time, such an equation was derived by H.C. Pocklington of England in 1897 and bears its name.

The differential operator in (9.2) can be subjected to *inversion* which is equivalent to solving (9.2) as a second-order differential equation in the function represented by the definite integral. On performing the necessary operations [1], we obtain what is known as the *integral Hallen equation* first analysed in detail by E. Hallen of Sweden in the 1930s-40s and, independently by M.A. Leontovich and M.L. Levin of the USSR, with respect to the unknown current distribution function on a dipole

$$\int_{z'=-l_2}^{l_1} \dot{I}_z(z')\,\dot{\mathscr{K}}(z-z')\,dz' = \dot{A}\cos\beta z + \dot{B}\sin\beta z - (2\pi\dot{V}/Z_m)\sin|\beta z| \tag{9.3}$$

where $\dot{V} = \dot{E}^d b$ = dipole-driving generator voltage

$\dot{A}$, $\dot{B}$ = arbitrary constants defined from the boundary conditions for the current at the dipole ends to turn to zero

$Z_m = (\mu_a/\varepsilon_a)^{1/2}$ = wave impedance of the medium surrounding the dipole

$\beta = \omega(\mu_a\varepsilon_a)^{1/2}$ = phase constant for a plane electromagnetic wave in the medium

An exact solution of Hallen's or Pocklington's integral equations is not known to have been obtained. That is the reason why in engineering calculations one most often uses a simplified *first-approximation solution*. This approximation is based on the fact that the equation kernel $\dot{\mathscr{K}}$ in the vicinity of the point $z = z'$ shows a well pronounced 'resonance' behaviour. Plots of functions used to express the real and imaginary parts of the kernel

$$\dot{\mathscr{K}}(z - z') = \beta(\cos\beta r/\beta r - j\sin\beta r/\beta r)$$

$$r = [(z - z')^2 + a^2]^{1/2}$$

are shown in Fig. 9.2. The plot for the real part of the kernel clearly shows that the vector potential of the electric current filament on the side surface of the dipole is mainly determined by the currents flowing near the point $z = z'$, and that for a/λ tending to zero the

effect of all other parts of the current filament on the vector potential may be neglected. Therefore, the integral on the left-hand side of Eq. (9.3) may be taken between the limits from $z' = z - h$ to $z' = z + h$, where h is some constant small in comparison with the wavelength. Then it may be taken that

$$\dot{\mathscr{K}} \approx [(z - z')^2 + a^2]^{-1/2}$$

and that the electric current within the interval of integration is constant and equal to the current at point $z' = z$. After the integral has been evaluated approximately, Eq. (9.3) reduces to the following approximate equality:

$$\dot{I}_z(z)\,\Omega \approx \dot{A}\cos\beta z + \dot{B}\sin\beta z - (2\pi\dot{V}/Z_m)\sin\beta\,|z| \quad (9.4)$$

where

$$\Omega \approx 2\pi \ln(2h/a)$$

is the result obtained by integrating the kernel over a small interval, $2h$, on placing the currents outside the integration sign.

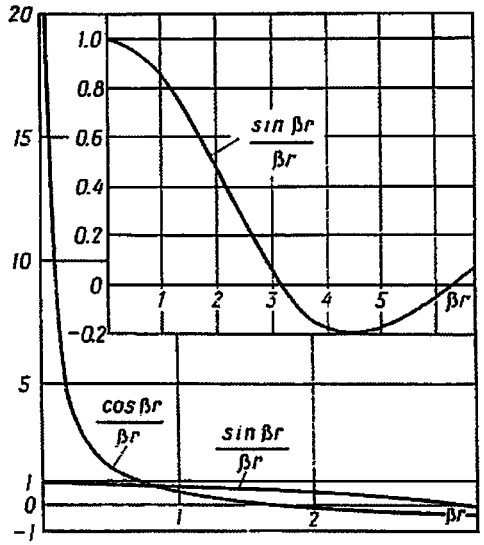


Fig. 9.2. Functions defining the kernel of the integral Hallen equation

Using the boundary conditions for the current at the dipole ends, $\dot{I}_z(-l_2) = 0$ and $\dot{I}_z(l_1) = 0$, we find the constants $\dot{A}$ and $\dot{B}$, substitute them in Eq. (9.4), and obtain the final expression for the *distribution of current on a thin dipole to a first approximation*

$$\dot{I}_z(z) = \dot{I}_0 \begin{cases} [\sin\beta(l_1 - z)]/\sin\beta l_1 & \text{for } z > b/2 \\ [\sin\beta(l_2 + z)]/\sin\beta l_2 & \text{for } z < -b/2 \end{cases} \quad (9.5)$$

where

$$\dot{I}_0 = (j4\pi\dot{V}\sin\beta l_1 \sin\beta l_2)/[Z_m\Omega\sin\beta(l_1 + l_2)]$$

is the current at the feed points (practically, at $z = 0$). For a *symmetrical dipole* in which the arms have the same length, $l = l_1 = l_2$*, the current distribution is symmetrical about the mid-point of the dipole:

$$\dot{I}_z(z) = [\dot{I}_0 \sin\beta(l - |z|)]/\sin\beta l \quad (9.6)$$

* In such cases, l is often referred to as the half-length of a dipole.—*Translator's note.*

Apart from the current distribution, it is of interest to know the electric-charge distribution along the dipole. This knowledge is important in, say, determining the maximum value of power at the dipole input that would cause a break-down of the surrounding medium. To a first approximation, the charge distribution can be found by the continuity equation, (A.3), which for a linear current takes the form

$$d\dot{I}_z(z)/dz + j\omega\dot{Q}(z) = 0 \tag{9.7}$$

where $\dot{Q}(z)$ is the charge per unit length of the dipole, in coulombs per metre. Applying (9.7) to (9.5) and noting that

$$\beta/\omega = (\mu_a \varepsilon_a)^{1/2} = 1/c$$

where c is the velocity of light in the medium surrounding the dipole, we obtain the following charge distribution

$$\dot{Q}(z) = (\dot{I}_0/jc) \begin{cases} [\cos\beta(l_1 - z)]/\sin\beta l_1 & \text{for } z > b/2 \\ [-\cos\beta(l_2 + z)]/\sin\beta l_2 & \text{for } z < -b/2 \end{cases} \tag{9.8}$$

Notably, for a symmetrical dipole,

$$\dot{Q}(z) = [\pm\dot{I}_0 \cos\beta(l - |z|)]/jc \sin\beta l$$

Here, the '+' sign applies to the positive values of z and the '—' sign, to its negative values.

Thus, in a *thin dipole the current and charge distributions vary as the circular sine*. In more detail, these distributions are shown in Fig. 9.3*a* and *b*. The plots show the current and charge distributions along symmetrical and unsymmetrical dipoles varying in electrical length. Of special interest is the most common *half-wave dipole* for which the total length is

$$l_1 + l_2 = \lambda/2$$

An important distinction of a half-wave dipole is the fact that its current and charge distribution functions do not depend on the location of the generator. In all other cases, the position of the driving generator has a decisive bearing on these distributions.

Let us formulate practical rules for determining the current and charge distributions for a thin dipole.

(1) At the dipole ends, the distribution has nodes (nulls) of current and antinodes of charge. Within a quarter-wavelength of the ends, the distribution has current loops and charge nodes (nulls). Within another quarter-wavelength, there appear current nodes and charge loops, and so on.

(2) At any point on dipole, the current and the charge are in time-phase quadrature (that is, shifted through 90 electrical degrees relative to each other).

(3) Both the current in and the charge on a dipole undergo a phase reversal (a phase change of 180 electrical degrees) at zero crossings.

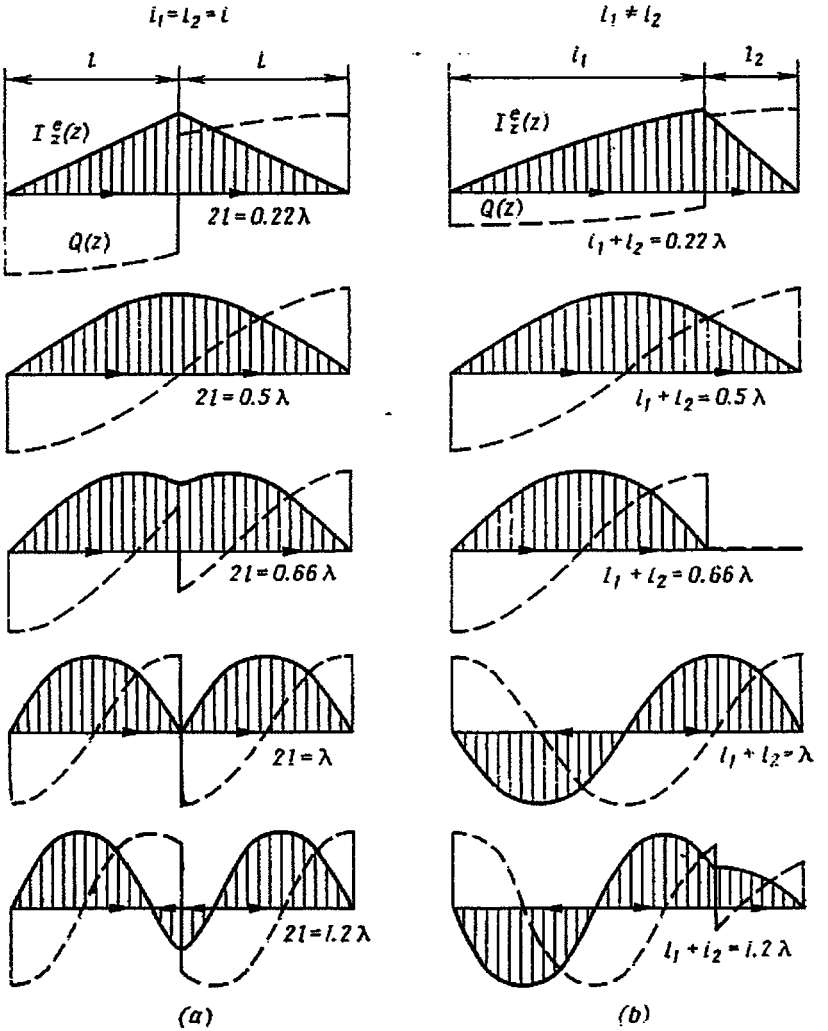


Fig. 9.3. Current and charge distribution in an electric dipole

(4) At the feed points, the current remains continuous and the charge changes stepwise.

(5) In an unsymmetrical dipole, current and charge loops are different in the different arms, and their ratio depends on the rela-

tive magnitude of l_1 and l_2:

$$I_1/I_2 = Q_1/Q_2 = \sin \beta l_2/\sin \beta l_1$$

Notably, at $l_1 \approx \lambda/2$ and $l_2 \neq \lambda/2$, the current and charge at any point on l_2 are close to zero ($l_1 + l_2 = 0.66\lambda$ in Fig. 9.3).

Note that the current and charge distributions just tend to vary as the sine when a/λ tends to zero, but they are never exactly sinusoidal. Equation (9.5) fails to hold especially near the current nodes where the vector potential is decided by the joint action of the currents flowing in sufficiently remote portions rather than by the local value of current at a given point on the dipole. The true current distribution at nodes cannot become zero and it grows increasingly more other than sinusoidal as the dipole increases in diameter.

Because to a first approximation the vector potential is mainly decided by the local value of current at a given point on the dipole, it follows that *the current and charge distributions defined by (9.5) and (9.8) remain valid for curved dipoles*, including corner and folded dipoles. In such cases, the coordinate z has the significance of the distance along the axis of the curved conductor.

Although the sine current distributions in dipole antennas are approximate, they are useful in tackling the external problem which has as its objective to determine the radiation pattern, directivity and radiation resistance of a dipole. This is because these parameters are integral characteristics of the current distribution function, and small errors near distribution nodes do not give a marked contribution upon integration.

9.2. The Radiation Pattern, Radiation Resistance and Directivity of a Dipole

In order to simplify calculations, only a symmetrical (or balanced) dipole will be discussed here and elsewhere in the book. Let us use the general procedure for the calculation of radiation fields, set forth in Sec. 7.3. Suppose that the symmetrical dipole in question is oriented in the z-direction and its mid-point, that is, its feed point is located at the origin of spherical coordinates as shown in Fig. 9.4. The vector potential in the far-zone field will solely have the z-component which, according to (7.7) is given by

$$\dot{A}^e_{z,\infty} = \frac{\exp(-j\beta R)}{4\pi R} \int_{-l}^{l} \dot{I}_z(z) \exp(j\beta z \cos\theta)\, dz$$

where $z \cos \theta$ in the exponent of the exponential function under the integration sign is the path length difference for the rays drawn to the field observation point from the origin of coordinates and the

instantaneous integration point. On placing under the integral sign the sine current distribution (9.6), utilizing the obvious relation $\dot{A}_\theta = -\dot{A}_z \sin\theta$, and applying the first line from (7.6), we obtain

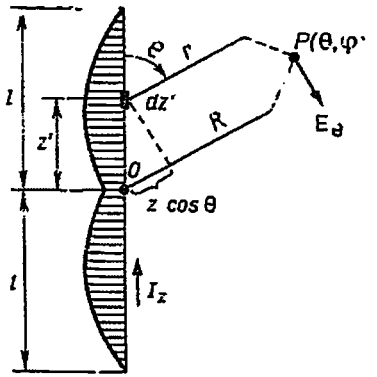


Fig. 9.4. To calculating the far field of a dipole

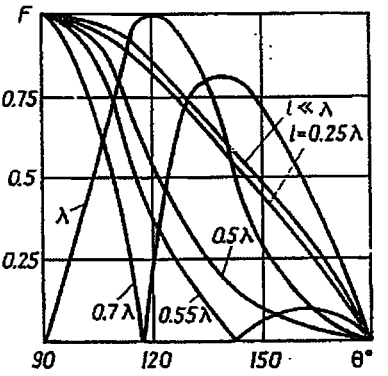


Fig. 9.5. Radiation patterns of a symmetrical dipole

an expression for the electric field strength (electric intensity) of the dipole

$$\dot{E}_\theta = \frac{j\dot{I}_l Z_m \sin\theta}{2\lambda}\left[\int_0^l \sin\beta(l-z)\exp(j\beta z\cos\theta)\,dz + \int_{-l}^{0} \sin\beta(l+z)\exp(j\beta z\cos\theta)\,dz\right]\exp(-j\beta R)/R$$

where the dipole input current $\dot{I}_0$ is replaced by $\dot{I}_l$, or the current at a distribution loop, which never turns to zero

$$\dot{I}_l = \dot{I}_0/\sin\beta l$$

The integrals can readily be evaluated by twice applying the rule for integration by parts:

$$\int \sin(ax+b)\exp(cx)\,dx = \frac{\exp(cx)}{a^2+c^2}[c\sin(ax+b) - a\cos(ax+b)]$$

The final expression for the far-zone electric field intensity is

$$\dot{E}_\theta = \frac{j\dot{I}_l Z_m}{2\pi}\,\frac{\cos(\beta l\cos\theta) - \cos\beta l\exp(-j\beta R)}{\sin\theta}\,\frac{}{R} \qquad (9.9)$$

Let us describe in brief the radiation field of a symmetrical dipole on the basis of what follows from analysis of Eq. (9.9).

1. The phase of the far-zone electric intensity is independent of observation angles. Therefore *the dipole has a phase centre* which coincides with the mid-point of the dipole.

2. The radiation of the dipole is independent of the azimuthal angle φ, which means that the radiation pattern in the equatorial plane (in the H-plane) is uniform.

3. The radiation field is zero along the axis of the dipole.

4. The shape of the radiation pattern in the meridional plane (the E-plane) depends on the ratio between the half-length of the dipole and the operating wavelength.

Figure 9.5 shows a family of typical radiation patterns in a meridional plane for dipoles varying in the ratio l/λ. The half-power beamwidth is 90° for l/λ tending to zero, 80° for $l/\lambda = 0.25$, and reduces to 44° for $l/\lambda = 0.55$. At l/λ in excess of 0.5, the radiation beam acquires side lobes. This is because with lengths like that, antiphase portions appear in the current distribution function along a dipole (Fig. 9.3). At $l/\lambda = 1$, the dipole ceases to radiate in the direction $\theta = 90°$, and its radiation pattern retains two identical lobes making an angle of $\pm 32°$ with the normal to the z-axis. Because of the break-up of the radiation pattern at half-lengths in excess of 0.55λ, very long dipoles are used extremely seldom.

The most commonly used form is a *half-wave dipole* with a half-length of 0.25λ for which the normalized radiation pattern takes, in agreement with Eq. (9.9), the following form

$$\dot{F}_\theta(\theta) = j \cos[(\pi/2)\cos\theta]/\sin\theta$$

For short symmetrical dipoles (with $\beta l \ll 1$), the cosines of the small arguments in Eq. (9.9) may be replaced by the first terms of the Taylor series:

$$\cos\alpha \approx 1 - \alpha^2/2$$

Then, on substituting them in (9.9), carrying out identity transformations, and noting that

$$\dot{I}_l = \dot{I}_0/\sin\beta l \approx \dot{I}_0/(\beta l)$$

we obtain the following expression for the radiation field:

$$\dot{E}_\theta = j\dot{I}_0 Z_m (l/2\lambda)\sin\theta \exp(-j\beta R)/R \qquad (9.10)$$

at $\beta l \ll 1$.

In Eq. (9.10) we go back to the current at the feed point because at $l/\lambda < 0.25$, there are no current loops on the dipole itself.

From comparison of Eq. (9.10) with Eq. (A.12a) for the electric field due to a Hertzian dipole it follows that a short symmetrical dipole with a sinusoidal (actually, a triangular) current distribution is equivalent in terms of the far field to half as long a Hertzian

dipole and has a similar normalized radiation pattern

$$F(\theta) = j \sin \theta$$

The *radiation resistance of a symmetrical dipole* is customarily referred to the current at a distribution loop, that is,

$$R_{\Sigma l} = \frac{2P_\Sigma}{|\dot{I}_l|^2} = \frac{1}{Z_m |\dot{I}_l|^2} \int_0^{2\pi} \int_0^{\pi} |\dot{E}_\theta|^2 R^2 \sin\theta \, d\theta \, d\varphi$$

On substituting here the expression for the field intensity of a dipole, Eq. (9.9), we obtain an equation for the radiation resistance at a loop

$$R_{\Sigma l} = (Z_m/2\pi) \int_0^{\pi} \frac{[\cos(\beta l \cos\theta) - \cos\beta l]^2}{\sin\theta} d\theta \qquad (9.11)$$

The integral in (9.11) might be expressed in terms of integral sines and cosines, but it appears more advantageous to carry out the integration directly on a computer. The integration result is shown

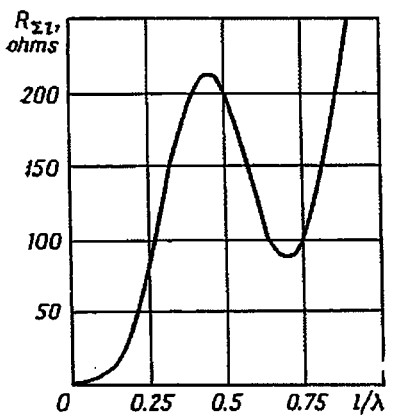


Fig. 9.6. Radiation resistance] of an electric dipole

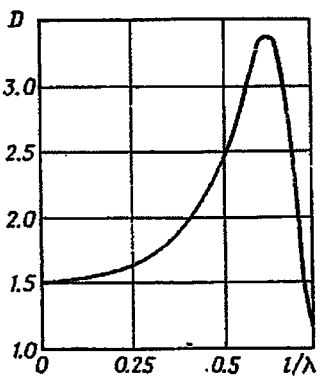


Fig. 9.7. Directivity of a symmetrical dipole in the direction $\theta = 90°$

in Fig. 9.6. It is useful to memorize that the radiation resistance of a half-wave dipole is 73.1 ohms and that of a full-wave dipole ($l = 0.5\lambda$) is 199 ohms. The oscillating behaviour of the $R_{\Sigma l}$ plot in Fig. 9.6 at $l > 0.5\lambda$ is due to the appearance of anti-phase portions in the current distribution function along the dipole. Now we may pass from the radiation resistance at a loop to that at the feed point, using the relations

$$P_\Sigma = 0.5 R_{\Sigma l} |\dot{I}_l|^2 = 0.5 R_{\Sigma 0} |\dot{I}_0|^2$$

that is,

$$R_{\Sigma 0} = R_{\Sigma l}/\sin^2 \beta l \qquad (9.12)$$

For example, in the case of short dipoles with βl tending to zero an approximate evaluation of the integral in (9.11) leads to the equation

$$R_{\Sigma l} = Z_m (\beta l)^4/6\pi$$

which, on referring the radiation resistance to the feed point, takes a more convenient form (noting that $\sin \beta l \approx \beta l$) as follows:

$$R_{\Sigma 0} = 20 (\beta l)^2 = 80\pi^2 (l/\lambda)^2$$

at $l/\lambda \ll 1$.

In determining the *directivity of a symmetrical dipole* in the direction $\theta = 90°$ (at $l/\lambda < 0.64$, this is the direction of maximum radiation), it is convenient to use Eq. (7.17) which takes the form

$$D = \Pi_{max}/\Pi_{av} = |\dot{E}_{\theta\ max}|^2 4\pi R^2/(Z_m |\dot{I}_l|^2 R_\Sigma)$$

On substituting here the field strength from (9.9) for $\theta = 90°$, we obtain an equation which is convenient to use when determining the directivity of a dipole, namely

$$D = Z_m (1 - \cos \beta l)^2/\pi R_{\Sigma l}$$

A plot of the directivity of a symmetrical dipole as a function of the ratio l/λ appears in Fig. 9.7. It is useful to note three typical values: $D = 1.64$ for a symmetrical half-wave dipole, $D = 2.41$ for a symmetrical full-wave dipole, and $D = 3.36$ for a dipole with $l = 0.625\lambda$. The fall in the directivity at l in excess of 0.625λ is due to changes in the shape of the radiation pattern, namely a decrease in the main lobe and an increase in the side lobes.

From the known values of directivity and radiation resistance it is an easy matter, using Eq. (7.21), to find the *effective length of a dipole, referred to the input current*

$$h_e = \lambda (DR_{\Sigma 0}/Z_m\pi)^{1/2} = \lambda (1 - \cos \beta l)/\pi \sin \beta l$$

As an alternative, the above expression could be derived by reducing the equation for the electric field strength, (9.9), to a form analogous to (7.11) by normalizing the radiation pattern in terms of the maximum value. It can be proved that the *effective length of a dipole is determined by the input current area of the same magnitude.* To demonstrate, if we integrate the sinusoidal distribution function (9.6) over the entire length of the dipole and refer the resultant integral to the current at the feed points, we will obtain

$$(2/\dot{I}_0) \int_0^l \dot{I}_0 \frac{\sin \beta (l-z)}{\sin \beta l} dz = (\lambda/\pi) \frac{1-\cos \beta l}{\sin \beta l} = h_e$$

Thus, the effective length of a dipole may be construed as the length of a hypothetical dipole with a uniform current distribution, which produces in the direction of maximum radiation the same field intensity as a real antenna. As an example, for a half-wave dipole, the effective length is 0.637 of its full length, and for a very short dipole (with βl substantially smaller than unity), it is equal to its half-length, that is, $h_e = l$.

9.3. Calculation of the Input Impedance of a Dipole by the Equivalent Circuit Method

The input impedance of a dipole is defined as the ratio of voltage to current at feed points, that is,

$$\dot{Z}_{in} = \dot{V}/\dot{I}_0$$

It has a vital bearing on whether or not the antenna input can be properly matched in the required frequency band.

There is no way of finding the input impedance of a dipole by solving the Hallen integral equation to a first approximation because the input impedance defined from the approximate equation (9.6) for the current distribution is, firstly, purely imaginary (which means that the radiation is not taken into consideration) and, secondly, indeterminate due to an arbitrary choice of the integration interval in deriving the intermediate equation (9.4). A need therefore arises to resort either to numerical methods for solving the integral equation (9.3) or (9.2) on a computer or to various approximations based on physical intuition.

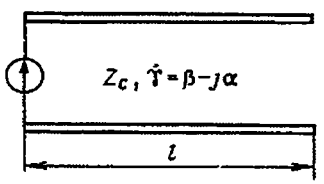


Fig. 9.8. Equivalent circuit of an electric dipole

Engineering practice has proved the worth of what is known as the *equivalent circuit method* which may well be extended to other antenna types. By this method the antenna under study is represented as a distributed-parameter or lumped-parameter equivalent circuit. The circuit element values are chosen such that the input impedance of the equivalent circuit could approximate the input impedance of the prototype antenna over the desired bandwidth in the best possible manner and could give a correct idea about the dependence of $\dot{Z}_{in}$ on antenna dimensions.

For a symmetrical dipole, it has proved advantageous to use an equivalent circuit in the form of an open-circuited, two-wire lossy transmission-line section (Fig. 9.8). The parameters of this circuit are the section length l, its characteristic impedance Z_c, and the

complex propagation constant

$$\dot{\gamma} = \beta' - j\alpha$$

where $\beta' = 2\pi\xi/\lambda$ = equivalent phase constant
ξ = correction factor
$\alpha = R_1/Z_c$ = equivalent attenuation constant
R_1 = resistance per unit length of one line conductor

The current distribution in the equivalent open-circuited line section varies as the hyperbolic sine and does not turn to zero at current nodes any longer.

The conditions of equivalence may be summed up as follows:

(1) The transmission-line length is assumed to be equal to the half-length (the length of one arm) of the dipole, l.

(2) The total power lost in the equivalent circuit and the radiated power of the dipole are deemed to be equal at all frequencies.

(3) The characteristic impedance of the equivalent circuit for a dipole of half-length l and of conductor radius a is given by the following approximate equation:

$$Z_c = (Z_m/\pi)\,[\ln(l/a) - 1] \quad (9.13)$$

which resembles in structure the equation for the characteristic impedance of a two-wire transmission line (the half-length of a dipole plays the role of the averaged spacing between the line conductors).

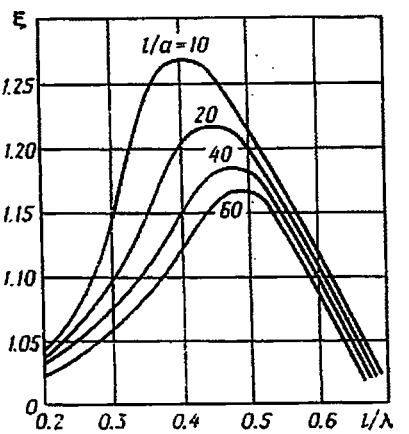


Fig. 9.9. Correction factor ξ

(4) The equivalent phase constant is chosen to be greater than the wave number in the medium surrounding the dipole by the correction factor ξ obtained from experimental data (Fig. 9.9). In effect, the factor ξ accounts for the capacitance of the dipole ends and also for the fact that actually the dipole is not a homogeneous line with uniformly distributed per-unit-length parameters.

Let us dwell in more detail on the second condition of equivalence which may be written as

$$0.5\,|\dot{I}_l|^2 R_{\Sigma l} = 2\int_{z=0}^{l} 0.5\,|\dot{I}_z(z)|^2 R_1\,dz$$

where the integral on the right-hand side is the total dissipated power in a two-wire transmission line for which the per-unit-length resistance is R_1. On placing under the integral sign the sinusoidal

current distribution on the dipole, and taking into account the relative phase velocity, we obtain

$$R_{\Sigma l} = 2R_1 \int_{z=0}^{l} \sin^2 [\xi\beta (l - z)]\, dz$$

On re-arranging the expression under the integral sign with the aid of the identity

$$\sin^2\alpha = 0.5 (1 - \cos 2\alpha)$$

and taking the integral, we obtain the desired design equation for the distributed radiation resistance per unit length of a symmetrical dipole

$$R_1 = \frac{R_{\Sigma l}}{l [1 - \sin (2\xi\beta l)/2\xi\beta l]}$$

Since the distributed radiation resistance is large in value, it is important to take into account its effect on the characteristic impedance of the dipole

$$\dot{Z}_{c1} = \left(\frac{2R_1 + j\omega L_1}{j\omega C_1}\right)^{1/2} = (L_1/C_1)^{1/2} (1 - j2R_1/\omega L_1)^{1/2}$$
$$\approx Z_c (1 - jR_1/\omega L_1)$$

where L_1 and C_1 is the inductance and capacitance per unit length of the equivalent line.

Noting that

$$R_1/\omega L_1 = \alpha Z_c/\omega L_1 = \alpha/\omega (L_1 C_1)^{1/2} = \alpha/\beta$$

we get

$$\dot{Z}_{c1} \approx Z_c (1 - j\alpha/\beta)$$

That is, we have found that the equivalent characteristic impedance $\dot{Z}_{c1}$ is a complex quantity.

Lastly, using equations from the theory of lossy long lines, we obtain the design equation for the input impedance of a dipole

$$\dot{Z}_{in} \approx Z_c (1 - j\alpha/\beta) \coth (\alpha l + j\xi\beta l)$$

where the characteristic impedance is as defined by Eq. (9.13), the attenuation constant is $\alpha = R_1/Z_c$, and the correction factor ξ is taken from experimental plots (see Fig. 9.9).

Fig. 9.10 shows plots of the resistive and reactive parts of the input impedance of a symmetrical dipole as a function of the ratio l/λ for three values of conductor diameter. Analysis of these plots shows the following.

(1) As the ratio l/λ varies from zero to 0.6, the input impedance of the dipole has two resonance regions. The first series resonance occurs at around $l/\lambda \approx 0.25$ and the second parallel resonance at l/λ

somewhat smaller than 0.5. At $l/\lambda < 0.25$, the reactive part of the input impedance is negative. As the conductor increases in diameter, the resonance value of the ratio l/λ decreases, especially so for parallel resonance.

(2) The input impedance of a dipole depends on frequency increasingly less as the dipole grows in diameter. To put this differently, the equivalent Q-factor of a dipole at frequencies of both the first

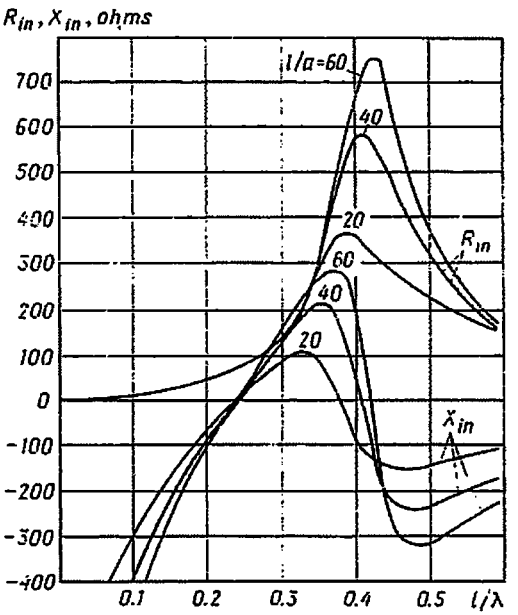


Fig. 9.10. Input resistance of an electric dipole

and second resonance decreases with decreasing characteristic impedance of the dipole. This behaviour can be explained as follows. The Q-factor of a circuit is proportional to the ratio between the stored electromagnetic energy and the energy lost over a cycle of r.f. oscillations at the resonance frequency. In the case of a dipole, some energy is lost by radiation such that its flow is decided solely by the ratio l/λ and does not practically depend on conductor diameter. In contrast, the stored energy is concentrated right near the dipole conductors, and its amount increases in inverse proportion to conductor diameter. A thick dipole stores less energy because the electromagnetic field is 'crowded' by the metal from the region of maximum concentration. Thus, *where it is desired to obtain a broad bandwidth, it is advantageous to use 'thick' dipoles so as to equalize the input impedance at the various frequencies.*

Note that the input impedance of a dipole antenna is strongly affected by the physical arrangement of the feed point — a factor entirely ignored in the design equation. The same goes for most antennas of other types. Therefore in practical antenna design the calculation value of input impedance is usually regarded as an approximate and tentative one and is usually verified and refined when carrying out experiments on a prototype antenna.

9.4. The Symmetrical Magnetic Dipole. Slot Antennas in an Infinite Plane Screen

Let us consider a cylindrical dipole of length $2l$ and of radius a, made in the form of an ideal magnetic conductor and symmetrically driven at the mid-point by an mmf $\dot{H}_z^d(z)$ which acts in a gap of width b (Fig. 9.11). Since the vector potential of magnetic currents obeys the same equation as the vector potential of electric currents and the boundary conditions as regards the magnetic field in the case of a magnetic dipole are the same as their counterparts for the electric field in the case of an electric dipole, it follows that in a symmetrical magnetic dipole the magnetic current distribution must satisfy the same integral (Hallen or Pocklington) equation as applies to a symmetrical electric dipole. In consequence, as a tends to zero, the magnetic current distribution in the arms (half-lengths) of a symmetrical magnetic dipole is to a first approximation a sinusoidal one

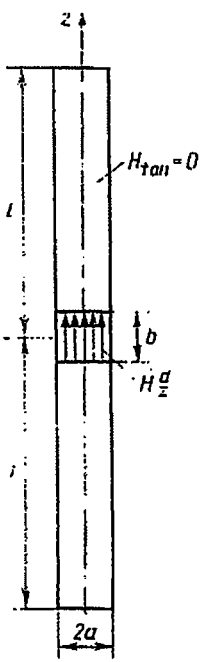


Fig. 9.11. Symmetrical magnetic dipole

$$\dot{I}_z^m(z) = \dot{I}_l^m \sin\beta\,(l - |z|) \qquad (9.14)$$

where $\dot{I}_l^m$ is the magnetic current at the maximum of the sinusoid, usually referred to as the magnetic loop current.

The replacement of $\dot{I}^e$ by $-\dot{I}^m$, the interchangeability of $\dot{\mathbf{E}}$ and $\dot{\mathbf{H}}$, and the replacement of Z_m by $-1/Z_m$, which all stem from the principle of permutational duality (see the Appendix), offers an opportunity to use Eq. (9.9) derived for the far-zone field of an electric dipole in order to immediately derive expressions for the far-zone electromagnetic field of a symmetrical magnetic dipole as follows:

$$\dot{H}_\theta = \frac{j\dot{I}_l^m}{2\pi Z_m} \frac{\cos(\beta l\cos\theta) - \cos\beta l}{\sin\theta} \frac{\exp(-j\beta R)}{R} \qquad (9.15)$$

and

$$\dot{E}_\varphi = -\dot{H}_\theta Z_m$$

and to establish connection between the radiation conductance $G^m_{\Sigma l}$ and the input admittance G^m_{in} of a magnetic dipole, on the one hand, and the radiation resistance $R^e_{\Sigma l}$ and the input impedance $\dot{Z}^e_{in}$ of an electrical dipole, on the other (it is assumed that the two dipoles are the same in shape and in size):

$$G^m_{\Sigma l} = R^e_{\Sigma l}/Z^2_{in}$$

and

$$\dot{G}^m_{in} = \dot{Z}^e_{in}/Z^2_{in}$$

Hence it follows that the input admittance of a magnetic dipole shows the same frequency dependence as the input impedance of a similar electrical dipole. Notably, while a half-wave electrical

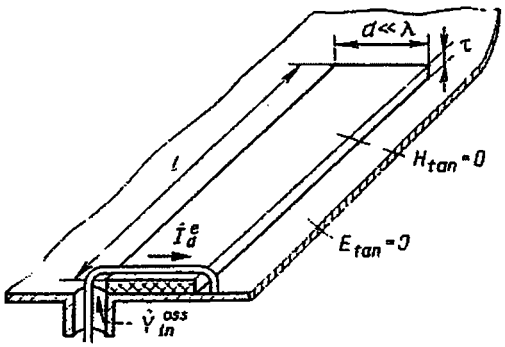


Fig. 9.12. Electrodynamic model of a one-sided slot antenna

dipole is, in the vicinity of its resonance frequency, equivalent to a lossy series resonant circuit, a half-wave magnetic dipole will, in the vicinity of its resonance frequency, be equivalent to a parallel resonant circuit.

Now let us analyse the properties of narrow slot antennas in plane screens, using the results obtained for a magnetic dipole. Suppose that we have a magnetic dipole which is rectangular in cross-section, d wide across the longer sides and τ wide across the shorter sides, such that $d \ll \lambda$. We arrange this magnetic slot dipole on the surface of an ideally conducting plane of infinite size. This leaves us with an electrodynamic model of a *one-sided slot* in a screen, excited at the centre by a conductor carrying an electric current $\dot{I}^e_d$ (Fig. 9.12). The slot is $2l$ long and d wide. Obviously, the distribution of the magnetic current (that is, electric field strength) in such a model of

the slot can be described by Eq. (9.14). Using the image method, we find that the electromagnetic field in the upper half-space into which the slot is radiating is twice the magnitude of the field set up by the same magnetic dipole in free space, and this means that this field is determined by twice the value given by Eq. (9.15). In the lower, or shadow, half-space of the screen the field is zero everywhere.

The radiation conductance and input admittance of a one-sided slot are likewise twice as great

$$G_{\Sigma l}^{o l s} = 2R_{\Sigma l}^{e}/Z_m^2$$

and (9.16)

$$\dot{Y}_{in}^{o l s} = 2\dot{Z}_{in}^{e}/Z_m^2$$

where $R_{\Sigma l}^{e}$ and $\dot{Z}_{in}^{e}$ are the radiation resistance and input impedance of an electric dipole of cross-sectional dimensions $d \times \tau$, placed in the same medium into which the slot antenna is radiating, while $G_{\Sigma l}^{oss}$ and $\dot{Y}_{in}^{oss}$ are the radiation conductance and input admittance of a one-sided slot in an infinite screen, of the same dimensions as the electric dipole.

As has been shown by M.A. Leontovich, the cross-sectional area of a metal strip of width $d \ll \lambda$ is equivalent, in the calculation of input impedances (admittances), to a circular cross-section of radius $a = d/4$. In view of this finding, when calculating the input admittance of a slot antenna we may use the equations from Sec. 9.3, applicable to a cylindrical electric dipole.

In addition to doubling the input admittance, a one-sided slot in a plane infinite screen has twice as great a directivity in comparison with an equivalent dipole in free space. The point is that the infinite screen causes the far-zone Poynting vector of a one-sided slot to quadruple although the radiated power only doubles as compared with a magnetic dipole in free space. For example, the directivity of a narrow, half-wave antenna with a one-sided slot in an infinite plane screen is equal to 3.28.

In real one-sided slot antennas the screening of radiation into the lower half-space is effected by cavity resonators or waveguides. Two factors are important here. Firstly, a resonator changes the distribution of the driving mmf along the slot in comparison with the case where the excitation is concentrated at the centre, and this may cause the voltage distribution in the slot to be different from a sinusoidal one. Secondly, a resonator has an intrinsic susceptibility which is lumped with the input admittance of the slot and brings about a change in the overall input admittance of the antenna.

In practice, use is most often made of *half-wave resonant slots* in which the voltage distribution does not depend on the distribution

of the driving mmf. On the other hand, the dependence of the overall input admittance on the resonator size of a slot antenna and the manner of resonator excitation can additionally be utilized to match the antenna input. (In terms of its equivalent circuit, a half-wave slot antenna together with its resonator takes, in the vicinity of its resonance frequency, the form of a parallel resonant circuit with an adjustable tapping-down factor relative to the input line.)

Next in popularity to one-sided, slot antennas are *two-sided slot radiators* in a thin metal sheet of large dimensions (Fig. 9.13*a*). An

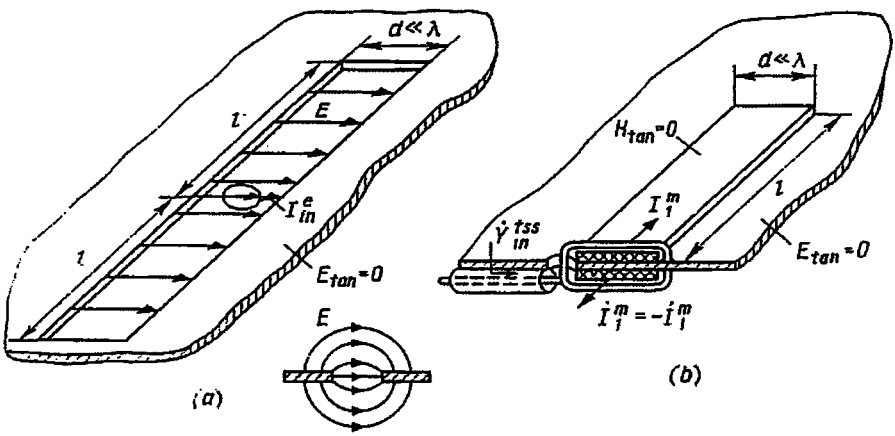


Fig. 9.13. Two-sided s ot in a sheet (*a*) and its electrodynamic model (*b*)

electrodynamic model of such an antenna may be a pair of strip magnetic dipoles arranged exactly one above the other on either side of an infinite metal sheet and driven by in-phase electric-current half-turns connected in parallel to a common generator (Fig. 9.13*b*). As before, the components of the electric field vector of a two-sided slot may be defined by Eqs. (9.15) multiplied by 2 in order to take into account the image effect. In this case, however, the radiation conductance and the input admittance of a two-sided slot are twice as great as the respective conductance and admittance of a one-sided slot due to radiation into both half-spaces:

$$G^{trs}_{\Sigma l} = 4R^e_{\Sigma l}/Z^2_m$$

and (9.17)

$$\dot{Y}^{tss}_{in} = 4\dot{Z}^e_{in}/Z^2_m$$

where $R^e_{\Sigma l}$ and $\dot{Z}^e_{in}$ are the radiation resistance and the input impedance of a metal strip dipole of the same dimensions as the slot. On the other hand, the directivity of a two-sided slot is just half as great as that of a one-sided slot, which means that it is the same as the directivity of a metal strip dipole in free space.

9.5. Coupled Dipoles. The EMF Method

Where it is desired to obtain a one-sided radiation and a concommittant increase in directivity, resort is made to antennas each consisting of several dipoles. Such an antenna will, in the general case, include *active* (or *driven*) *dipoles* whose inputs are each connected to a driving generator, and *passive* (or *parasitic*) *dipoles* which are not connected to any generators and which are excited by the electromagnetic field of the active dipoles. The inputs of the passive dipoles are usually terminated in load reactances which will not cause a reduction in the efficiency of the antenna.

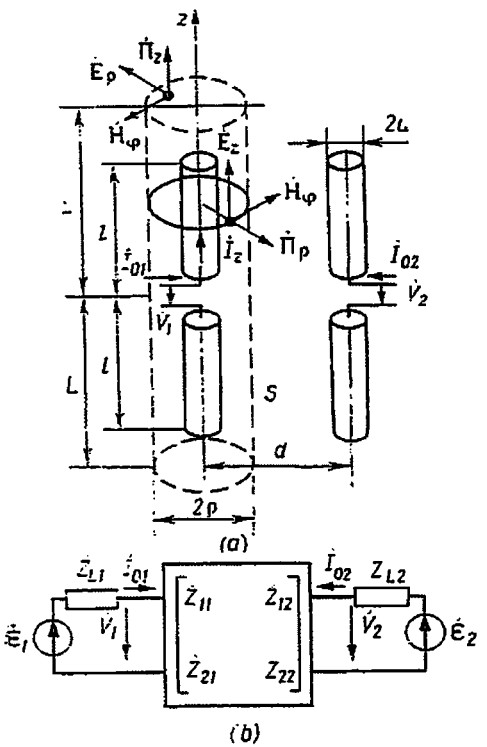


Fig. 9.14. Coupled dipoles: *a*, general arrangement; *b*, equivalent circuit

The simplest antenna system made up of two identical, parallel dipoles is shown in Fig. 9.14*a*. In order to determine the currents and voltages at the dipole inputs, it is convenient to depict this system as an equivalent two-port characterized by an impedance matrix **Z** (Fig. 9.14*b*). The diagonal elements of the matrix, $\dot{Z}_{11}$ and $\dot{Z}_{22}$, are the *self-impedances* of the dipoles, determined for one dipole while the other dipole is open-circuited. The off-diagonal elements, $\dot{Z}_{12} = \dot{Z}_{21}$, are the *mutual impedances*, taking care of the electromagnetic coupling between the two dipoles. Preference is given to the **Z** matrix because the dipoles open-circuited for its determination are considered not to contribute to the radiation field (this assumption is satisfied more accurately in the case of resonant half-wave and shorter dipoles). Therefore, it may be expected that the addition of further dipoles to the system will not change the self and mutual impedances that existed prior to the inclusion of still another dipole.

When the spacing between the thin dipoles of an antenna is not too small (that is, with $d \gg a$ and $a \ll l$), it seems legitimate to a first approximation to write the sinusoidal current dis-

tribution along the dipoles

$$\dot{I}_{z1,2}(z) = \dot{I}_{01,2}[\sin\beta(l_{1,2} - |\dot{z}|)]/\sin\beta l_{1,2} \qquad (9.18)$$

where the coordinate z is reckoned from the mid-point of each dipole. The input currents $\dot{I}_{01}$ and $\dot{I}_{02}$ should be found by simultaneously solving the linear Kirchhoff equations

$$\left.\begin{aligned}\dot{V}_1 &= \dot{I}_{01}\dot{Z}_{11} + \dot{I}_{02}\dot{Z}_{12} = \dot{\mathscr{E}}_1 - \dot{I}_{01}\dot{Z}_{L1}\\ \dot{V}_2 &= \dot{I}_{01}\dot{Z}_{12} + \dot{I}_{02}\dot{Z}_{22} = \dot{\mathscr{E}}_2 - \dot{I}_{02}\dot{Z}_{L2}\end{aligned}\right\} \qquad (9.19)$$

where $\dot{\mathscr{E}}_1$, $\dot{\mathscr{E}}_2$ = emf of the driving generators

$\dot{Z}_{L1}$, $\dot{Z}_{L2}$ = internal impedances of the generators

For a passive dipole, it should be taken that $\dot{\mathscr{E}}_i = 0$ and that $\dot{Z}_{L\,i}$ is the load impedance.

The self and mutual impedances of dipoles can conveniently be found by what is known as the *emf method* first proposed in 1922 independently by D.A. Rozhansky of the USSR and L. Brillouin of France and later improved by I.G. Kliatskin, V.V. Tatarinov and A.A. Pistolkors of the USSR. The emf method boils down to the following. Surround one of the vibrators (say, the first one) by a closed cylindrical surface S (Fig. 9.14*a*) of height $2L$ and of radius ρ. The Poynting vector components normal to this surface can be written in the local coordinate system of the dipole as

$$\dot{\Pi}_\rho = -0.5\dot{E}_z\dot{H}^*_\varphi$$

and

$$\dot{\Pi}_z = 0.5\dot{E}_\rho\dot{H}^*_\varphi$$

Obviously, the integral of the normal component of the Poynting vector over the surface of the cylinder is the power drawn by the dipole from the generator. Since near the dipole the vectors $\dot{\mathbf{E}}$ and $\dot{\mathbf{H}}$ are not in phase (as they are in the far-zone field), the power turns out to be a complex quantity, with its active part in a lossless medium being equal to the power radiated by the dipole.

Let us arrange the surface S to lie in the plane of the dipole surface so that $\rho = a$ and $L = l$. Then with the dipole radius a tending to zero, the integral of the Poynting vector z-component, $\dot{\Pi}_z$, taken over the end surfaces of the cylinder will tend to zero, and the power will be determined by the integral taken solely over the side surface

$$\dot{P}_1 = \int_{-l}^{l}\int_{0}^{2\pi}\dot{\Pi}_\rho a\,dz\,d\varphi = -0.5\int_{-l}^{l}\dot{E}_{z1}(a)\times 2\pi a\dot{H}^*_{\varphi 1}\,dz$$

where account is taken of the fact that when a tends to zero and is substantially smaller than d, the field near the dipole does not practically depend on the azimuthal coordinate φ.

Let us write the total tangential component $\dot{E}_{z1}(a)$ on the side surface of the dipole as the sum

$$\dot{E}_{z1} = \dot{E}_{z11} + \dot{E}_{z12}$$

where $\dot{E}_{z11}$ is the component due to the currents in the first dipole and $\dot{E}_{z12}$ is the component due to the currents in the second dipole, and also note that

$$2\pi a \dot{H}^*_{\varphi} = \dot{I}^*_{z1}$$

is the complex conjugate current of the first dipole. Then the expression for $\dot{P}_1$ will take the form

$$\dot{P}_1 = -0.5\left(\int_{-l}^{l} \dot{E}_{z11}\dot{I}^*_{z1}\,dz + \int_{-l}^{l} \dot{E}_{z12}\dot{I}^*_{z1}\,dz\right)$$

Alternatively, the power $\dot{P}_1$ may be written in terms of the voltage and current at the dipole input. On taking into account Eq. (9.19), we get

$$\dot{P}_1 = 0.5\,\dot{I}^*_{01}\dot{V}_1 = 0.5\,(\dot{I}^*_{01}\dot{I}_{01}\dot{Z}_{11} + \dot{I}^*_{01}\dot{I}_{02}\dot{Z}_{12})$$

On comparing the two expressions for $\dot{P}_1$, we arrive at the key design equations used in the emf method for evaluating the mutual and self impedances of dipoles

$$\left.\begin{aligned} \dot{Z}_{12} &= -\frac{1}{\dot{I}^*_{01}\dot{I}_{02}}\int_{-l}^{l} \dot{E}_{z12}\dot{I}^*_{z1}\,dz \\ \text{and} \\ \dot{Z}_{11} &= -\frac{1}{|\dot{I}_{01}|^2}\int_{-l}^{l} \dot{E}_{z11}\dot{I}^*_{z1}\,dz \end{aligned}\right\} \qquad (9.20)$$

Expressions for $\dot{Z}_{22}$ and $\dot{Z}_{21} = \dot{Z}_{12}$ can be derived by simply interchanging the subscripts *1* and *2*.

Instead of the self and mutual impedances of dipoles referred to the dipole inputs, we may well use the *mutual and self impedances referred to loop currents on each dipole*, $\dot{Z}_{12l}$ and $\dot{Z}_{11l}$. The relations

involved are obvious and simple [see Eq. (9.12)]

$$\dot{Z}_{12l} = \dot{Z}_{12} \sin \beta l_1 \sin \beta l_2$$

and

$$\dot{Z}_{11l} = \dot{Z}_{11} \sin^2 \beta l_1$$

The tangential components $\dot{E}_{z12}$ and $\dot{E}_{z11}$ in (9.20) are in effect the emfs per unit length along the dipole surface, which are due to the currents in either the second or the first dipole, which is why the method is called the emf method. It is legitimate to ask: Why is it that on the side surface of the dipole the tangential components are non-zero and the integration in (9.20) is extended to include the entire length of the dipole instead of the expected region of the excitation gap where the tangential electric field should be concentrated? The answer is based on two points. Firstly, the tangential component vanishing on the side surface of the dipole is that of the *total* electric field and not of its individual parts. Secondly, and most importantly, instead of the unknown exact current distributions in dipoles the emf method uses the first-approximation sinusoidal distribution defined by (9.18), which cannot guarantee that the boundary conditions will be rigorously satisfied.

A major distinction of (9.20) as regards coupled dipoles is the fact that small errors in specifying the current distribution (first-order infinitesimals) produce substantially smaller errors in the mutual and self impedances of dipoles (second-order infinitesimals). That is why the emf method is widely used in the engineering design of dipole antennas.

The emf method gives the best results for the mutual impedances of half-wave resonant dipoles, where the current distribution does not depend on the location of the feed point and the distribution of the driving emf. On the other hand, the emf method is markedly inferior to the equivalent-circuit method as regards the self-impedances of coupled dipoles. The numerical implementation of the emf method requires knowledge of the near-zone electromagnetic fields, that is, the use of the exact operator $\hat{\mathscr{E}}(\dot{\mathbf{J}}^e)$, in agreement with Eq. (7.2). In the case of sinusoidal current distributions such as given by (9.18), the calculation of near-zone fields is markedly simplified and it is quite possible to develop relatively simple analytical expressions (see Secs. 2.5 and 2.6 in [1]) that form the basis of standard computer sub-programs.

The results obtained in evaluating the magnitudes and phases of the mutual impedances of parallel half-wave dipoles are shown in Fig. 9.15. In the limiting case of d tending to zero, that is, when the two dipoles are arranged to merge together, the mutual impedance

is transformed into the self-impedance (in ohms) of a thin half-wave dipole and takes on a value of $73.1+j42.5$, ohms. As d increases, the mutual impedance $\dot{Z}_{12}$ gradually decreases in magnitude, and its phase changes in an almost linear fashion such that an increase in d by one wavelength brings on an additional delay of 2π. The dashed lines in Fig. 9.15 show the mutual impedance of dipoles as found by the asymptotic equation (9.18) which in the case of parallel half-wave dipoles in air reduces to the expression

$$\dot{Z}_{12} = 60j\lambda \exp(-j\beta d)/\pi d$$

It can be seen that the asymptotic equation (9.40) yields fairly accurate results when the two dipoles are spaced apart by a distance greater than a wavelength.

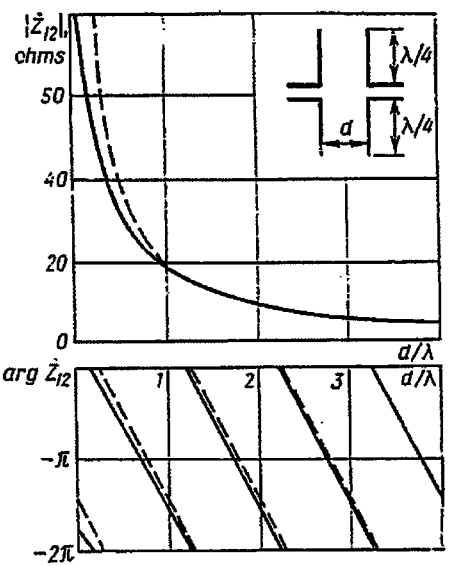


Fig. 9.15. Magnitude and phase of the mutual impedance of dipoles

By invoking the principle of interchangeability between electric and magnetic currents [see. Eqs. (A.8) in the Appendix], the emf method can readily be converted to the *mmf method* for thin magnetic dipoles and narrow slots in plane screens. In so doing, one replaces in Eqs. (9.19) $\dot{V}$ by $\dot{I}$, $\dot{I}$ by $\dot{V}$, and $\dot{Z}_{ik}$ by $\dot{Y}_{ik}$ so that the impedance matrix **Z** used to describe a system of coupled dipoles is replaced by the admittance matrix **Y**.

9.6. The Pattern Multiplication Theorem in the Analysis of the Radiation Pattern of a System of Identical Parallel Dipoles

By virtue of the linearity of Maxwell's equations, the electromagnetic field set up by a system of radiators (a so-called *antenna array*) is the sum of the fields due to the individual elements of the system. If these elements are the same in size, have the same distribution of radiating currents, and are oriented in space in the same manner (Fig. 9.16), *the far electromagnetic field can be represented as the product of two factors.* One of these, the *element factor*, is the field radiated by a single (reference) element, $\dot{\mathbf{F}}_{el}(\theta, \varphi)$. The other, the *array factor*, $\dot{f}_{\Sigma}(\theta, \varphi)$ is due to the system of isotropic point radi-

ators placed at the locations of the system elements. Thus,

$$\dot{\mathbf{E}}(R, \theta, \varphi) = \dot{A}\dot{\mathbf{F}}_{el}(\theta, \varphi)\dot{f}_{\Sigma}(\theta, \varphi)\exp(-j\beta r)/R \qquad (9.21)$$

with R tending to infinity. Here, $\dot{A}$ is the amplitude factor dependent on the total power of the coherent generators feeding the array. *The array factor* $\dot{f}_{\Sigma}(\theta, \varphi)$ *is sometimes called the pattern factor* or *the space factor.*

Equation (9.21) is a mathematical record of the *pattern multiplication theorem* or *method** which defines the manner in which the radiation patterns of a system of identical radiators are multiplied together in order to obtain the overall result. To prove it, we assign to each radiator a local coordinate system made up of parallel axes x_n, y_n, z_n, where n is the nth radiator. The origin of each local coordinate system is arranged to occur at some characteristic point, such as the phase (or radiation) centre of the respective radiator. In local coordinates, the vector radiation patterns of the radiators can be calculated by the technique formulated in Sec. 7.3 [see Eqs. (7.6) through (7.8)]. Naturally, in view of the assumptions made, these radiation patterns will be the same and can be described by a common function, $\dot{\mathbf{F}}_{el}(\theta, \varphi)$. In the principal coordinate system x, y, z, which remains parallel to the axes of the local coordinate systems, the radiation patterns of the individual radiators will, in accordance with Eq. (7.16), all differ as regards the directional phase characteristic by a quantity dependent on the angles of observation. This quantity is $\beta R_n \cos \alpha_n$, where R_n is the distance to the origin of the coordinate system of the nth radiator and α_n is the angle between the direction to the observation point in the far-zone field and the direction to the origin of the local coordinate system of the nth radiator (Fig. 9.16). Here, $\dot{J}_{ec}$ stands for an extraneous current.

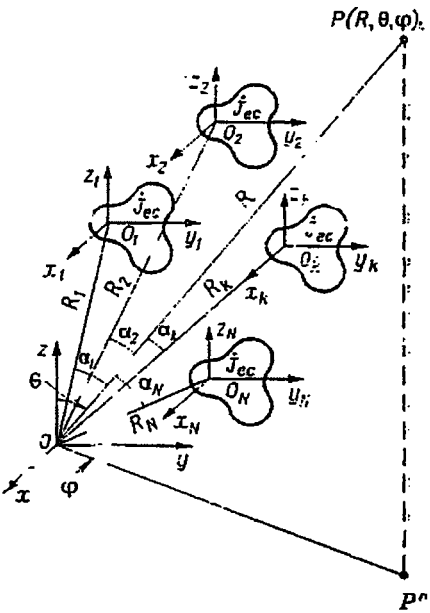


Fig. 9.16. System of identical radiation sources (antenna array)

* It is also known as the product pattern theorem. See "Reference Data for Radio Engineers", 5th Edition, p. 23-25.—*Translator's note.*

Since the complex excitation amplitudes of the individual radiators may be different, we obtain the following expression for the overall radiation pattern of a system containing N elements:

$$\dot{\mathbf{f}}(\theta, \varphi) = \sum_{n=1}^{N} \dot{I}_n \dot{\mathbf{F}}_{el} \exp(j\beta R_n \cos\alpha_n) = \dot{\mathbf{F}}_{el} \left[\sum_{n=1}^{N} \dot{I}_n \exp(j\beta R_n \cos\alpha_n) \right]$$

Thus, we have proved Eq. (9.21), and the scalar array factor in it has the form

$$\dot{f}_{\Sigma}(\theta, \varphi) = \sum_{n=1}^{N} \dot{I}_n \exp(j\beta R_n \cos\alpha_n) \tag{9.22}$$

where

$$R_n \cos\alpha_n = x_n \sin\theta \cos\varphi + y_n \sin\theta \sin\varphi + z_n \cos\theta$$

The pattern multiplication (or product pattern) theorem offers an opportunity to see which features of the overall radiation pattern of an antenna system owe their origin to the properties of a single element (this above all applies to the field polarization) and which features are due to the interference of the fields set up by the isotropic sources described by the array factor. For example, an electric dipole may be treated as a superposition of an infinite number of Hertzian dipoles spaced apart along the dipole arms.

Let us apply the pattern multiplication theorem to analysis of the radiation pattern of the system of two parallel dipoles shown in Fig. 9.14*a*. Assuming that the origin of the principal coordinate system lies at the mid-point of the first dipole and that the second dipole is at a distance d along the y-axis, we are in a position to write an expression for the array factor at once

$$\dot{f}_{\Sigma}(\theta, \varphi) = 1 + \dot{m} \exp(j\beta d \sin\theta \sin\varphi)$$

where $\dot{m} = \dot{I}_{02}/\dot{I}_{01} = m \exp(j\Delta\Phi)$ = ratio of input currents, definable from the solution of Eqs. (9.19)

φ = azimuthal angle

θ = meridional angle

The plot of the array factor is the shape of a body of revolution about the y-axis. The shapes of the section obtained by passing the xy-plane through the plot of the array factor, with the dipoles excited with equal-amplitude currents, with d equal to 0.25 λ, and for several phase shifts between the driving currents are shown in Fig. 9.17.

In the direction of the x-axis (where the path length difference is zero), in-phase excitation of the dipoles gives rise to a maximum of radiation whereas excitation with currents in anti-phase produces nulls (or zeros) of radiation. In the case of excitation with currents in quadrature (that is, when $\Delta\Phi = \pi/2$) and with $d = 0.25$ λ, the

plot of the array factor has the shape of a cardioid with a maximum pointing in the direction of the dipole carrying the current delayed in phase. In this case, the dipole reflecting the radiated energy towards the other dipole is called an *active reflector*. An active reflector doubles the directivity of a single dipole in the direction of maximum radiation.

If the other element is a *passive* (or parasitically) excited dipole, then, assuming that $\dot{\mathcal{E}}_2 = 0$ and $\dot{Z}_{L2} = jX_{L2}$, we can immediately write the expression for the current ratio as

$$\dot{m} = \dot{I}_{02}/\dot{I}_{01} = -\dot{Z}_{12}/(\dot{Z}_{22} + jX_{L2}) \qquad (9.23)$$

As follows from (9.23), with X_{L2} equal to zero, that is, when the input of the passive dipole is short-circuited, it will always carry

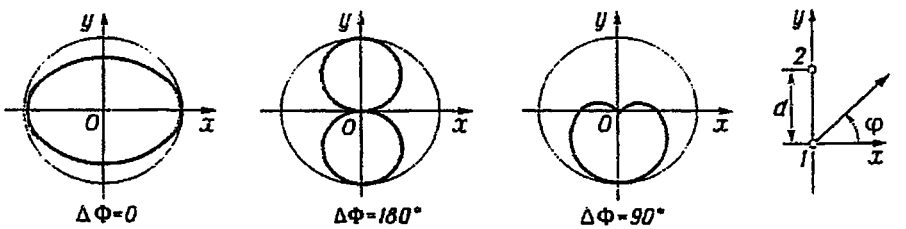


Fig. 9.17. Pattern factor of two equal-amplitude isotropic sources ($d/\lambda = 0.25$)

a smaller current than will the active dipole, and this current will decrease progressively with an increase in the spacing between the dipoles. The passive dipole can be tuned to resonance (when $X_{22} + X_{2L} = 0$) either by adjusting the load reactance while keeping the dipole length unchanged, or by varying the length of the dipole arms while short-circuiting the input. At resonance, the current in the passive dipole becomes maximal at any value of d, and if d is small, this current can even exceed that in the active dipole.

Of practical importance are two modes of operation for passive dipoles: the *reflector mode* with the radiation maximum pointing in the direction of the active dipole, and the *director mode* with the major lobe of radiation pointing towards the passive dipole. In each case, the spacing between the two dipoles is chosen to be equal from 0.15 to 0.25 of a wavelength. Only one degree of freedom exists for the choice of the relative values of currents in the dipoles—the value of the load reactance. Therefore, it is not feasible to avoid backward radiation (opposite to the major lobe of radiation) completely, and the plot of the array factor for a system incorporating a passive dipole is other than a cardioid diagram. As a rule, a passive dipole is tuned and aligned so as to minimize backward radiation as much as practicable.

The current in a reflector element should lead in time phase the current in the active dipole. Noting the behaviour of $\dot{Z}_{12}$ (see Fig. 9.15), analysis of Eq. (9.23) shows that for this requirement to be satisfied, the sum of the tuning impedance and the reactive part of the self-impedance should be negative. The current in a director element should lag in time phase behind the current in the active dipole. This is achieved with the aid of a tuning capacitive reactance. As often as not, the passive dipole is short-circuited, and the tuning is done by varying its length. *In the reflector mode, the length of the passive dipole should be somewhat in excess of* 0.5 λ, *whereas in the director mode, the figure should be somewhat below* 0.5 λ. How much the passive half-wave dipole should be made longer or shorter depends on the spacing between the two dipoles and on the conductor diameter, $2a$.

As a way of enhancing the directional properties of an array, it may, in addition to an active dipole, include a reflector and, at the same time, one or several directors. In Soviet practice, such a system is called a *director-type antenna*. More commonly, it is referred as the Uda-Yagi, Yagi-Uda, or, simply, Yagi antenna.

Chapter Ten

Low-Directional and Frequency-Independent Antennas

10.1. Physical Arrangement of Dipole Antennas

The simplest and most commonly used type of antenna is the *half-wave dipole*. When a two-wire transmission line is connected to such a dipole (Fig. 10.1*a*), the SWR of the line is 5 to 7, and this

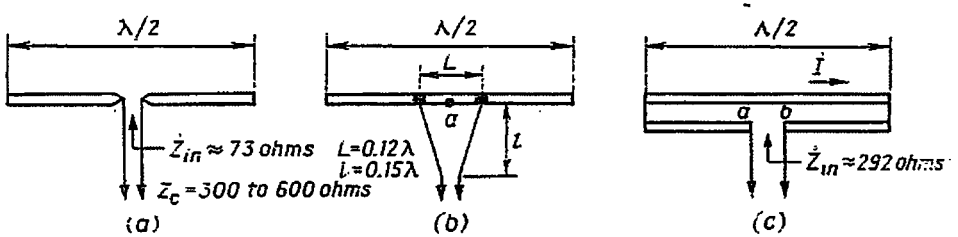


Fig. 10.1. Half-wave dipoles

requires that the line to be tuned to sustain a travelling-wave condition. The dipole turns out to have a limited bandwidth. It is convenient to use a shunt-fed half-wave dipole arranged as shown in

Fig. 10.1*b*. By varying the values of L and l of the shunt, it is possible to obtain on the line a condition close to the travelling-wave mode without resort to any additional elements. Another advantage of shunt feed is the fact that a voltage node is produced at the point a of the dipole, and the dipole may thus be secured to its support without insulators.

A higher input impedance can be obtained with a *folded-dipole antenna* (known as the Pistolkors type in Soviet practice). As is seen from Fig. 10.1*c*, it comprises two parallel, closely spaced dipole antennas both of which are connected together at their ends by

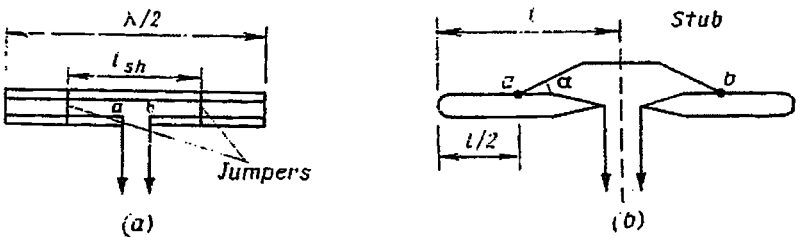


Fig. 10.2. Shunt dipoles

jumpers, and one is fed at its midpoint over a two-wire transmission line. The parallel elements of the dipole carry in-phase currents with maxima at the middle and, since the spacing between the conductors is small in comparison with a wavelength, the radiation pattern of a folded-dipole antenna is the same as that of a simple half-wave dipole.

The input impedance of a folded-dipole antenna can be calculated on the basis of the power balance. Let the antinode current in each dipole be I. Then the total radiated power will be

$$P_{\Sigma} = (2I)^2 R_{\Sigma l}$$

On the other hand, this power may be expressed in terms of the input impedance as

$$P_{\Sigma} = I^2 R_{ab}$$

Hence, the input impedance of a folded dipole is

$$R_{ab} = 4R_{\Sigma l}$$

At $R_{\Sigma l} = 73.1$ ohms, we have R_{ab} approximately equal to 292 ohms, which is close to the characteristic impedance of two-wire lines and there is no need to insert a separate matching device.

A folded dipole can be fitted with auxiliary jumpers as shown in Fig. 10.2*a*. In such a case, the length of the shunt, l_{sh}, and the diameters of the conductors that make up the dipole can be varied so as to obtain a good match between the input impedance of the

antenna and the characteristic impedance of a two-wire transmission line over a rather wide frequency band.

Still another variation of a shunted symmetrical dipole is shown in Fig. 10.2*b*. With the proper choice of shunt location, shunt size, the angle α, and the characteristic impedance of the line, it is possible to obtain a good match over a frequency band with a four-fold ratio of maximum to minimum frequency (from $\lambda = 6.5\ l$ to $\lambda = 1.5\ l$) at an SWR of not over 3.3. Half-wave symmetrical dipoles

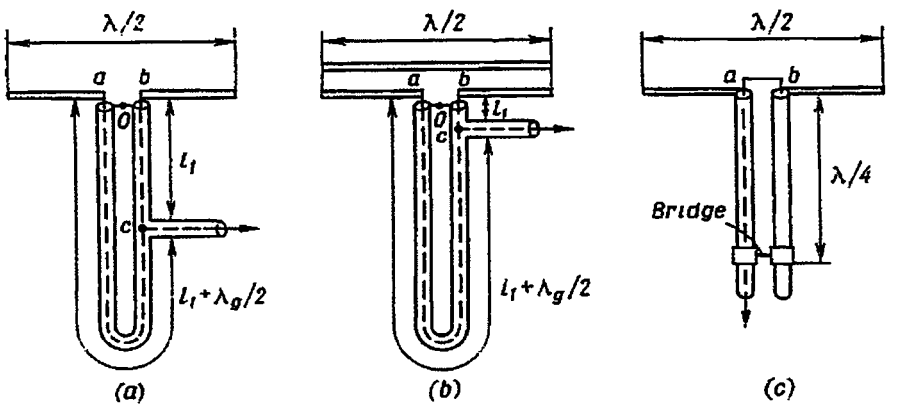


Fig. 10.3. Feeding of dipoles over a coaxial cable

can also be fed by a flexible coaxial cable or a rigid coaxial waveguide. In such cases, a need arises to use balancing devices, such as an *U-bend balun* (see Fig. 1.19).

In the arrangement of Fig. 10.3*a*, the input impedance of the dipole is about 73 ohms, and the resistances that load the cable sections are $R_{a0} = R_{b0} \approx 36.5$ ohms. If the cable has a characteristic impedance of $Z_c = 75$ ohms, and the length of the cable section is $l_1 = \lambda_g/4$, the resistance loading the feeder cable at point c will be

$$R_c = Z_c^2/2R_{a0} \approx 77 \text{ ohms}$$

and the SWR in the feeder cable of the antenna will be close to unity.

The arrangement of Fig. 10.3*b* involves a folded dipole with an input impedance of around 292 ohms. Therefore, the resistances that load the cable sections are the same and equal to $R_{a0} = R_{b0} \approx$ ≈ 146 ohms. If the length l_1 be taken equal to zero, the resistance loading the feeder cable at point c will be halved, that is, equal to around 73 ohms — a figure ensuring a condition close to a perfect match.

Figure 10.3*c* shows a *bridge-type balancing device* having a greater bandwidth. Here, the dipole is connected to two tubes fitted with a shorting bridge and making up, between them, a quarter-wave metallic insulator. One tube encloses a coaxial cable with its outer conductor connected to the tube at point *a*, while its inner conductor is connected to the other tube at point *b*. This arrangement avoids excitation at the outer surface of the cable and assures a balanced feed to the two halves of the dipole. In the case of a coaxial cable with a characteristic impedance of 75 ohms, a nearly travelling-wave condition is maintained on it. The metallic insulator plays

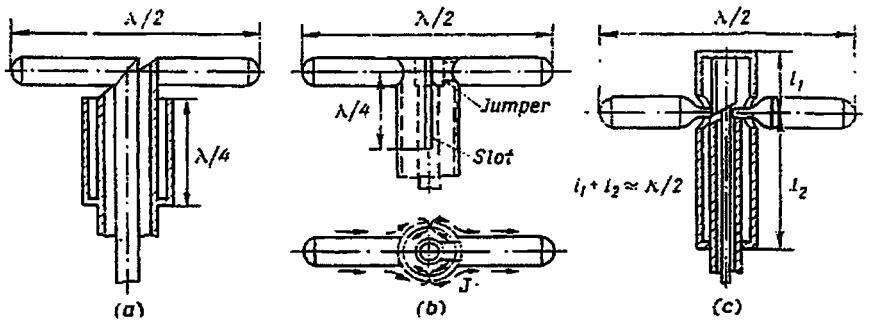


Fig. 10.4. Feeding of dipoles over a rigid coaxial line

the part of a shunt stub which balances out the reactive component of the input impedance and extends the bandwidth of match. The balancing stub may be enclosed in an additional shield so as to avoid parasitic radiation and to protect against external influences.

In the SHF band, half-wave symmetrical dipoles are driven by rigid coaxial waveguides. Here, too, a need exists for balancing devices that would prevent current leaking out to the external surface of the conductor in the coaxial line. In the arrangement of Fig. 10.4*a*, there is a coaxial waveguide over which is slipped a quarter-wave tube shorted to the outer conductor of the coaxial line at the lower end. This device, known as a *quarter-wave skirt*, acts as a metallic insulator — it serves to prevent some of the currents to flow onto the outer surface of the coaxial waveguide and to ensure a balanced feed to the dipole branches.

Another popular arrangement shown in Fig. 10.4*b* uses a symmetrical dipole which is energized via two longitudinal slots cut in the outer conductor of a coaxial waveguide. At the location of the slots, the inner conductor of the waveguide is joined to the outer conductor by a jumper. The jumper shorts out the TEM wave of the coaxial waveguide, and the current flowing in it excites higher modes, above all the TE_{11} wave, in the coaxial waveguide. Although this mode

cannot be propagated, its appearance is accompanied by local transverse electrical currents. These currents cross the slots in the outer conductor, thus exciting the external space. In consequence, transverse currents also appear on the outer surface of the coaxial guide and drive the branches of the symmetrical dipole. The dipole is excited most strongly when the slots have a resonant length (around a quarter of a wavelength).

The manner in which a symmetrical dipole is fed by means of a balancing and matching device is shown in Fig. 10.4*c*. Here, the coaxial guide is enclosed in a tube of length $l_1 + l_2$, shorted at both

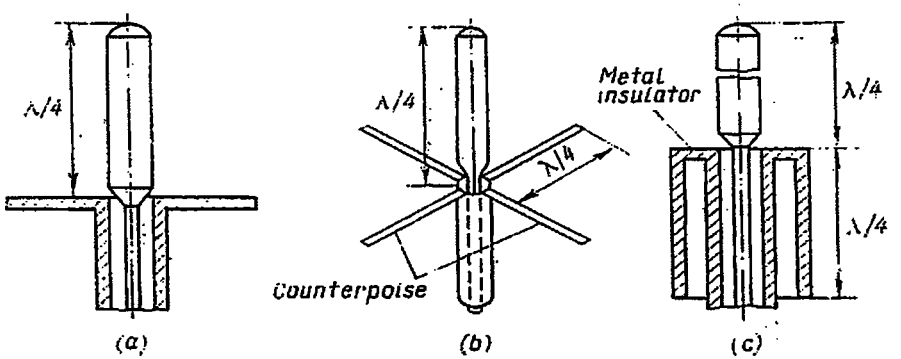


Fig. 10.5. Feeding of unsymmetrical dipoles

ends. The tube has openings through which the branches of a half-wave dipole are passed. One branch is joined to the end of the outer conductor in the coaxial waveguide, and the other, to the inner conductor of the guide which increases in thickness over the length l_1 and is joined to a tube. The overall length of the twin balancing coaxial stub is

$$l_1 + l_2 \approx 0.5\,\lambda$$

and the branch length ratio, l_1/l_2, is chosen so as to obtain the broadest possible bandwidth of match.

Now we turn to *unsymmetrical* (or *unbalanced*) *dipoles* fed by coaxial lines and used alone or set up on the conducting surfaces of various objects. Figure 10.5*a* shows a vertical quarter-wave dipole excited by a conducting disc to which the outer conductor of a coaxial line is connected. The disc may be replaced by four (or more) horizontal wires (Fig. 10.5*b*) quarter-wave long each. In another arrangement, the disc or the wires acting as counterpoises may be replaced by a quarter-wave skirt shorted at one end, as shown in Fig. 10.5*c*. Here, the outer surface of the skirt is excited so that it combines with the protruding inner conductor of the coaxial line to form a half-wave dipole.

On conducting surfaces, resort may be made to quarter-wave folded dipoles, notably those incorporating a shunt (Fig. 10.6*a*). Dipoles incorporating a shunt have an increased equivalent thick-

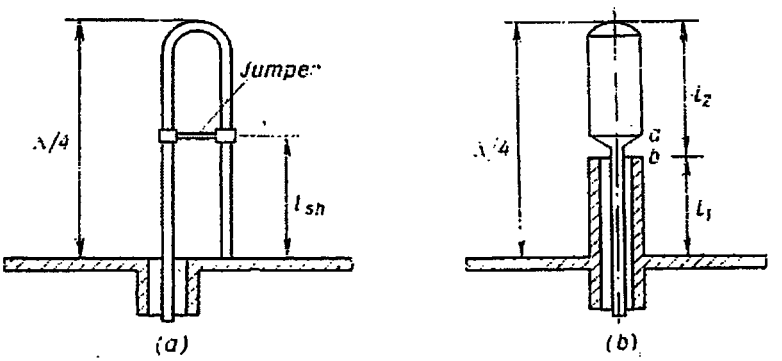


Fig. 10.6. Unsymmetrical dipoles with an extended bandwidth

ness in comparison with a simple dipole and have therefore an extended bandwidth.

Figure 10.6*b* shows the arrangement of a widely used quarter-wave dipole with a *rised feed point*. Here, the capacitive reactance of the upper half of the dipole at feed points *a* and *b* is balanced out by the series-connected inductive reactance of the lower half of the dipole. By varying the dipole diameter and the ratio between the arm lengths l_1 and l_2, it is possible to obtain a bandwidth of 10 to 15% for the antenna.

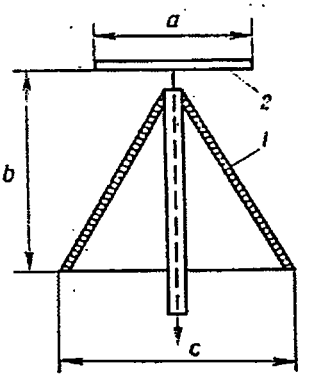


Fig. 10.7. Discone antenna

A still greater bandwidth is provided by what is called the *discone antenna* (Fig. 10.7) driven by a coaxial line. The line shield is connected to a metal cone, *1*, and the inner conductor to a metal disc, *2*. Beginning at some maximum wavelength, λ_{max}, the SWR of the line turns out to be close to unity, and the quality of match improves with decreasing wavelength. The antenna dimensions found by experiment are as follows: $a/\lambda_{max} = 0.175$, $b/\lambda_{max} = 0.25$, $c/\lambda_{max} = 0.275$. In the range of wavelengths from $a/\lambda = 0.175$ to $a/\lambda = 0.45$, the vertical radiation pattern changes in shape very little, being very close to that of the usual dipole.

10.2. Turnstile Antennas

The name '*turnstile*' is applied to an antenna made up of two dipoles normal to each other and arranged so that their axes intersect at their midpoints (Fig. 10.8).

A turnstile antenna can operate in any one of several modes, depending on the relative magnitude of the currents at the dipole inputs.

Omnidirectional (also known as all-directional or non-directional) operation. When the dipoles of a turnstile antenna are energized

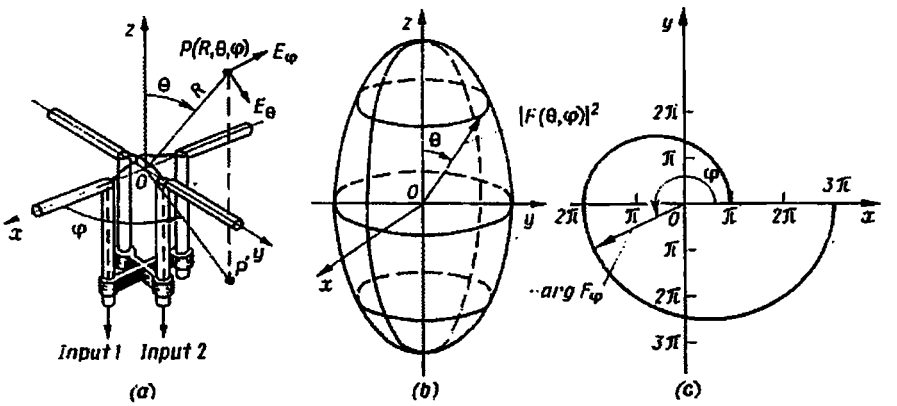


Fig. 10.8. Turnstile dipole: *a*, construction; *b*, field-strength radiating pattern; *c*, phase radiation pattern

by currents of equal amplitude and shifted through $\pi/2$ in phase, the spatial field-strength radiation pattern has the shape shown in Fig. 10.8*b*. There are *no nulls* (or *zeros*) *of radiation*, and the maxima occur in directions normal to both dipoles, that is, along the positive and negative *z*-directions. In the plane containing the dipoles, the radiation is linearly polarized, while in all other directions it is elliptically polarized. In the *z*-direction, the polarization of the radiation is purely circular. In the two half-spaces separated by the plane *xOy*, the polarization vector rotates in opposite directions. Thus, the *radiation field contains all likely states of polarization.* Of particular interest is the directional phase characteristics in the *xOy* plane (Fig. 10.8*c*) — they look like Archimede's spirals. The fact that the equi-phase lines do not close on themselves in moving round the antenna in the *xOy* plane in the far-zone field is an indication that *there is no phase centre.*

Omnidirectional radiation is produced by horizontally polarized transmitting TV antennas set up on tall masts. Along with the usual cylindrical dipoles, it is convenient in such antennas to employ what are known as *batwing dipoles* first proposed by B.V. Braude of the

Soviet Union (Fig. 10.9). The batwing element is a number of parallel metal tubes taking up the space within the enclosing rectangular frame. The dipole arms are shorted to the mast (for lightning protection) at points b-b, and the excitation voltage is fed at points a-a. The optimal dimensions of a batwing element are as follows: $a = 0.26\,\lambda_{av}$, and $l = 0.16\,\lambda_{av}$. The input impedance is then nearly constant and equal to 140 ohms over a bandwidth of $\pm$ 7.5%.

The vertical radiation pattern of a TV transmitting antenna can be reduced through the use of several in-phase energized tiers of

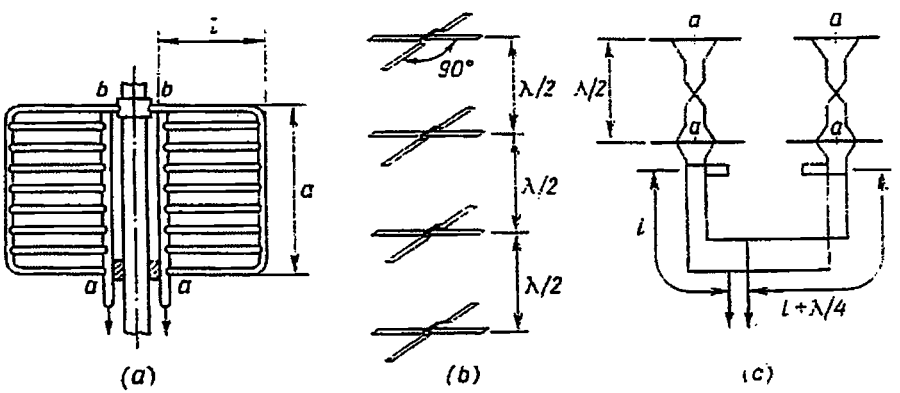


Fig. 10.9. Batwing dipole (*a*) for a multi-tier turnstile (superturnstile) TV transmitting antenna (*b*) with a two-wire feed line (*c*)

turnstile dipoles spaced about a half-wavelength apart as shown in Fig. 10.9*b*. With an even number of tiers, there is no radiation either upwards or downwards, and the major lobe of radiation is produced in a horizontal plane. The dipoles belonging to different tiers but lying in the same vertical plane are energized from one two-wire line, and the dipoles lying in the other vertical plane (normal to the first) are energized from another two-wire line (Fig. 10.9*c*). The lines are tuned to operate under travelling-wave conditions and are connected in parallel, one being a quarter of a wavelength longer than the other—this gives the desired phase difference of $\pi/2$ between the dipoles in each tier. The two-wire lines 'cross' between the tiers, thus ensuring that the dipoles in the different tiers are energized in-phase.

Controlled polarization operation. Obviously, by adjusting the amplitudes and phases of the currents fed to the crossed dipoles in the z-direction (that is, in the direction of the major lobe of radiation), it is possible to obtain any desired state of polarization. This offers an added opportunity to *effect polarization selectivity* in a radio system. In transmission, the currents fed to the dipoles can be

controlled with the aid of electronically controlled phase shifters and power dividers (see Secs. 4.8, 6.6, and 6.9). In reception, the dipoles of a turnstile antenna can be connected to separate receivers, and the signals from each radio channel can then be combined at the intermediate frequency or even at the audio frequency (that is, after detection). In the latter case, the receiving antenna and the associated receiver can receive waves of any polarization.

10.3. Slot Antennas

The fact that slot antennas can be made flush with a metal surface makes them extremely convenient for use on flying craft where a low aerodynamic drag is an important asset. Radiating slots cut in an infinite metal sheet can be analysed with relative ease by invoking the duality principle and the image method. In practice, slots

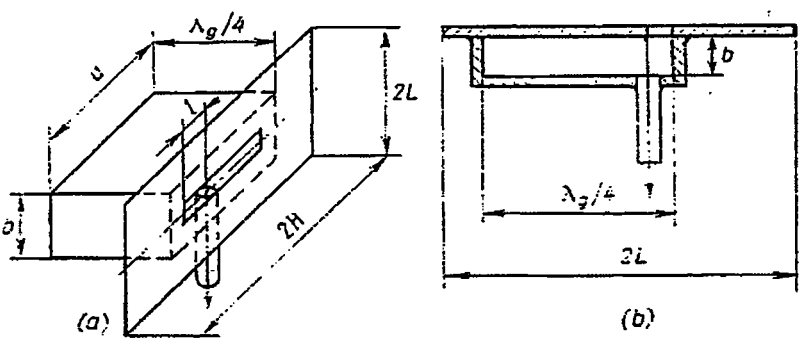


Fig. 10.10. Excitation of a slot antenna with a resonator (a 'boxed-in' slot)

are made in bounded surfaces of an irregular shape, and this necessitates the use of approximate techniques in order to calculate radiation characteristics. The simplest approach consists in that the curvilinear metal surface is replaced by a tangential metal sheet—an infinite flange. It has been found that this approach is good to calculate the radiation admittance of resonant slots on surfaces for which the linear dimensions and radii of curvature are at least equal to a wavelength.

The simplest half-wave slot antenna on a metal sheet of finite dimensions can be energized by a resonator in the form of a shorted quarter-wave rectangular waveguide segment operating in the TE_{10} mode (Fig. 10.10*a*). The antenna is energized by a coaxial cable in which the outer conductor is connected to one end of the slot, and the inner conductor, to the opposite end. The resonator acts as a metal insulator, and its susceptance balances out the intrinsic susceptance of the slot.

As follows from Eq. (9.16), the input impedance of a half-wave slot antenna incorporating a tuned resonator in a screen of large electrical dimensions, located at a voltage maximum (at the centre of the slot), is

$$R_s = Z_m^2/2R_{\Sigma l}^e = 970 \text{ ohms}$$

In order to match this impedance to the characteristic impedance Z_c of the coaxial line, the line should be connected within a distance l of the slot edge, as found from the equation

$$\sin^2\beta l = Z_c/R_s$$

Physically, it is convenient to locate the resonator so that one of its broad faces is adjacent to the screen and the other broad face passes the coaxial line inside the resonator, as shown in Fig. 10.10*b*.

Let us dwell on the radiation pattern of slot antennas constructed in a screen of finite dimensions. In the **H** vector plane (that is, in the plane passing through the slot), the radiation along the slot is zero, the diffraction of waves by the screen edge tells only slightly, and the radiation pattern does not depend on the dimensions of the screen. In the front half-space it is about the same as the radiation pattern of a slot in an infinite screen, and there is a minor lobe of radiation in the backward half-space.

In the **E** vector plane (that is, the plane normal to the slot), the radiation pattern strongly depends on the screen dimension $2L$ and only slightly depends on the screen dimension $2H$. The diffraction of waves by the screen edge is strongly manifested in this plane. The excitation of the screen edges gives rise to an additional radiation which is superimposed on the own radiation of the slot. This has, as a result, two consequences. Firstly, a field appears beyond the shadow side of the screen; secondly, the shape of the radiation pattern in the front half-space is distorted. As the waves radiated by the slot itself interfere with those radiated due to the diffraction by the screen edges, the radiation pattern displays maxima and minima whose number increases with increasing ratio $2L/\lambda$.

The radiation pattern of a slot antenna in the E-plane, as determined by experiment, is shown in Fig. 10.11 by the dashed curves. Shown in the same figure as full curves is the theoretical radiation pattern of the same antenna, as found by an approximate technique in which the distribution of the surface electric current on the front side of the screen is assumed to be the same as the distribution of the surface electric current on an infinite screen (this is the so-called *physical optics approximation*). Figure 10.12 shows the theoretical radiation pattern in the H-vector plane, as found under the assumption that the dimension $2H$ is finite and the dimension $2L$ is infinite. The screen in this case is an infinite strip which is a degenerate

elliptical cylinder. The theoretical calculations in this case have been carried out by the rigorous method of eigen-functions.

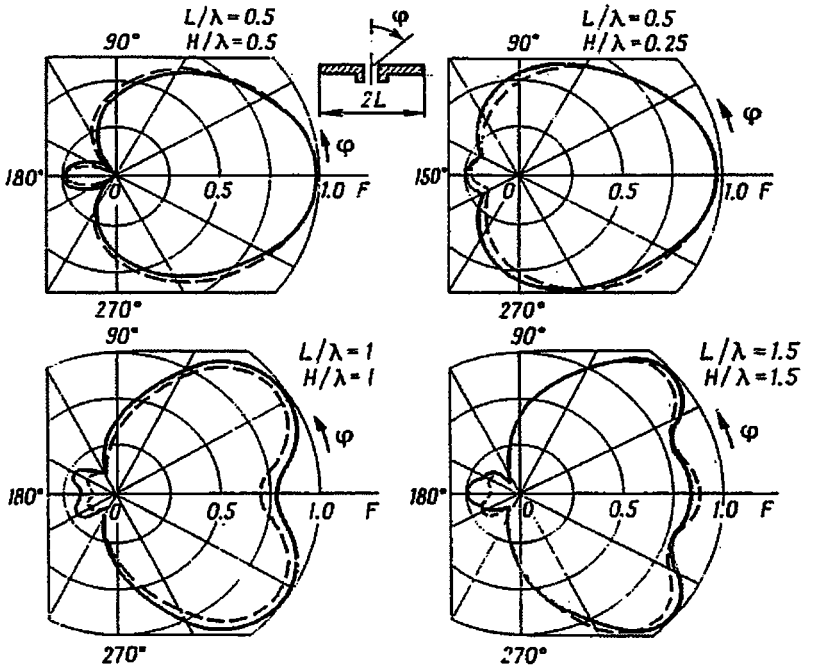


Fig. 10.11. Radiation patterns of a slot antenna in a finite screen in the E-vector plane

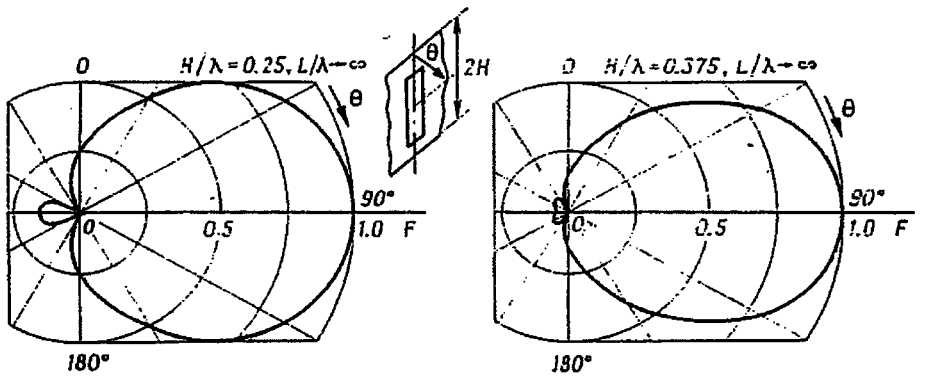


Fig. 10.12. Radiation patterns of a slot antenna in a finite screen in the H-vector plane

In practice, use is made not only of one-sided, but also of *two-sided slots* made in sheets of a limited size. If in the antenna of

Fig. 10.10*a* we remove the resonator, we will obtain a two-sided slot antenna. In both half-spaces the radiation patterns are about the same as in Figs. 10.11 and 10.12 for the illuminated region of space. However in the E-plane the radiation along the screen surface is exactly equal to zero. The input impedance of a two-sided slot antenna in a screen large in comparison with a wavelength can be found by Eqs. (9.17). For a resonant half-wave slot antenna it is about 485 ohms. Experimental plots of the resistive and reactive parts of

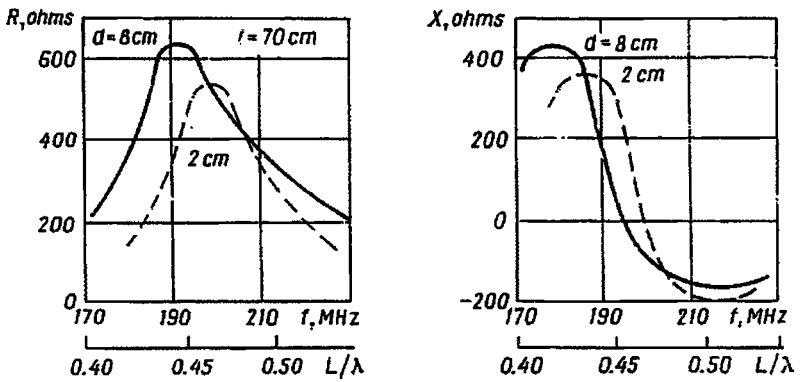


Fig. 10.13. Experimental plots of the input impedance of a two-sided slot ('boxed-in' slot) antenna

the input impedance near resonance for a two-sided slot antenna in a large screen are shown in Fig. 10.13. As is seen, the resonant slot length L is somewhat shorter than a half-wave-length and decreases with increasing slot width d.

In the SHF band, wide use is made of *annular-slot antennas*, with slots cut in a body of revolution, notably a circular metal cylinder, as shown at *1* in Fig. 10.14*a*. The annular slot, *4*, is made with the aid of a coaxial cavity, *2*, in the form of a quarter-wave metal skirt shorted over the cylinder. The cavity is energized by a coaxial line, *3*, at point *a*. If the average perimeter of the cavity is such that

$$\pi (r_1 + r_2) < \lambda$$

only a TEM wave can exist and will be most vigorously excited in it. With the cavity tuned to resonance, the slot voltage will be uniformly distributed in azimuth, with longitudinal electrical current flowing at the surface of the cylinder, so that the resultant antenna is not unlike a cylindrical electric dipole. In the azimuthal plane, the radiation will be uniform, while in the meridional plane the radiation pattern will substantially depend on the length of the driven cylinder.

If, on the other hand,

$$\pi (r_1 + r_2) > \lambda$$

the TEM mode excited in the cavity will be accompanied by the TE_{11} mode of the coaxial line for which the cutoff wavelength is

$$\lambda_{co} \approx \pi (r_1 + r_2)$$

Let the cavity be a half-wavelength long. Then the TEM-mode will be suppressed, but the TE_{11} mode will persist. The slot voltage will then be sinusoidally distributed in the azimuthal direction, whereas

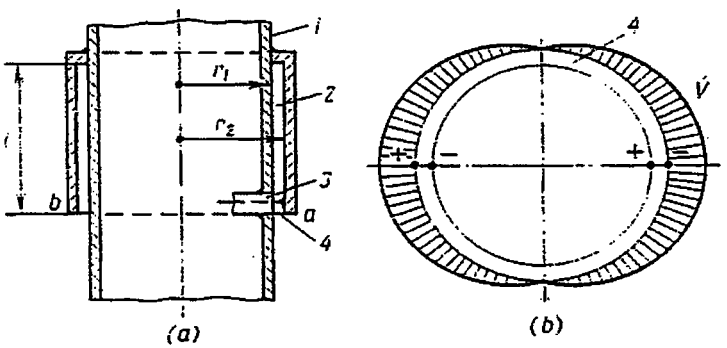


Fig. 10.14. Annular slot antenna in a cylinder

in addition to longitudinal currents, there will appear the transverse components of surface electric currents, and the radiation pattern of the antenna will be changed appreciably. The transverse components of surface currents will give rise to radiation along the axis of the cylinder. The antenna will radiate in all directions, and the radiation pattern will only slightly depend on the cylinder length.

Figure 10.14*b* shows the slot voltage distribution when the antenna is energized at two diametrically opposite points on the cavity. The slot voltage at point *a* is in anti-phase with that at point *b*. Therefore, the TEM mode is suppressed with any length l of the cavity, and the slot voltage is distributed in azimuth sinusoidally.

Of special interest is the *annular-slot antenna with a slot voltage wave travelling in azimuth*. In this case, the cavity (Fig. 10.14) is energized at four equidistant points on the slot, with every next point shifted in phase from the previous one through $-\pi/2$. If the cavity length is chosen equal to $\lambda_g/4$ for the TE_{11} wave, the slot voltage (neglecting any higher modes) will be varying in a travelling-wave fashion

$$\dot{V}(\varphi) = \dot{V}_0 \exp(-j\varphi)$$

The travelling-voltage-wave annular-slot antenna radiates along the z-axis a circularly polarized electromagnetic field similar to that radiated by a turnstile antenna (see Fig. 10.8). As has been shown by calculations and experiment, the radiation pattern of such an antenna only slightly depends on the size and shape of the metal body in which it is made.

10.4. Strip and Microstrip Antennas

Similarly to other microwave devices (power dividers, phase shifters, and matching devices), strip and microstrip antennas are fabricated by the integrated-circuit technology. Among their merits

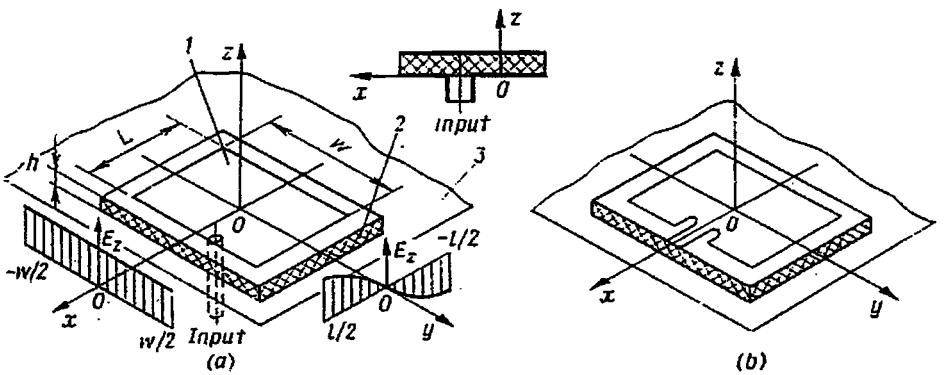


Fig. 10.15. Rectangular microstrip antennas fed by (*a*) a coaxial line and (*b*) by a strip line

are simplicity, small size and weight, and low profile or, which is the same, low aerodynamic drag (which fact is of special value for flying craft).

The simplest strip radiator (Fig. 10.15) consists essentially of *a metal sheet 1*, a *ground plane 3*, and a *dielectric substrate 2*. The metal sheet or patch may widely differ in shape. Most commonly, it may be rectangular, circular, or elliptical. It can be energized either by a coaxial line via a slot in the ground plane and substrate (Fig. 10.15*a*) or by a strip line lying in the plane of the radiating sheet (Fig. 10.15*b*). In the former case, the shield of the coaxial cable is connected to the ground plane, and the central conductor is passed through the substrate and connected to the metal sheet at a specially chosen point. In the latter case, excitation is conveyed to the feed point by an unbalanced (or unshielded) strip line for which purpose a cut-out is made in the radiating sheet (there may be no cut-out if the feed point is shifted towards an edge of the radiating sheet). The substrate is most often one of the dielectrics

for which $\varepsilon_r \approx 2.5$ to 10.0 and the loss tangent is anywhere between 10^{-3} and 10^{-4}. The substrate thickness h may be anywhere between one-hundredth and one-tenth of a wavelength.

Microstrip antennas fall in the class of *resonant antennas*. In simplified terms, it may be deemed that the resonant cavity of a microstrip radiator is bounded by vertical walls made of an ideal, magnetically conducting material and arranged all the way round the perimeter of the radiating sheet. In rectangular microstrip antennas use is ordinarily made of the lowest type of resonance for which

$$L \approx \lambda_g/2$$

where

$$\lambda_g \approx \lambda_0 (\varepsilon_r)^{1/2}$$

is the wavelength in the strip line with a conductor whose width w is less than, or equal to, a half of the wavelength. The electric field component $\dot{E}_z$ between the radiating sheet and the ground plane is distributed nearly uniformly across the strip line (in the y-direction) and sinusoidally (with antinodes at the sheet edges) longitudinally (in the x-direction). Marked departures from this idealized distribution are only observed nearly the corners of the radiating sheet and next to the feed point.

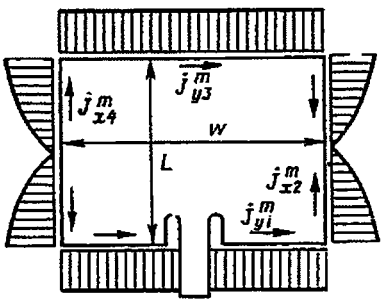


Fig. 10.16. Magnetic current distribution in a slot radiator equivalent to a microstrip antenna

The radiating system of a microstrip antenna may be modelled in more than one way. For example, one may take into consideration the surface electric currents flowing over the radiating sheet and the ground plane, and also the vertical currents of electrical polarization in the dielectric. However, it is a formidable electrodynamic task to establish the distribution of radiating electric currents, the more so that the region where these currents spread out markedly exceeds the size of the sheet. That is the reason why such a model is used but seldom, and then only when one seeks to take into account the finite dimensions of the ground plane by the method of physical optics. More often a simpler model is used for the radiating system, in which the *microstrip antenna is treated as an equivalent slot antenna in an infinite plane screen without a dielectric*. Studies have shown that this model is quite accurate in giving insight into the operating principle of the antenna and in carrying out tentative calculations of the radiation field. It is assumed that the equivalent slot has the same shape as the edges of the metal radiating sheet.

Figure 10.16 shows the distribution of magnetic currents $\dot{J}_x^m$ and $\dot{J}_y^m$ in the equivalent slot. It is constructed on the basis of the $\dot{E}_z$ field distribution in the cavity of a rectangular microstrip antenna. The principal role in shaping the radiated field is played by the uniformly distributed in-phase currents $\dot{J}_{y1}^m$ and $\dot{J}_{y3}^m$ which produce a linearly polarized radiation whose vector E is parallel with the x-axis. The currents $\dot{J}_{x2}^m$ and $\dot{J}_{x4}^m$ contain on each lateral side of the sheet two anti-phase sections whose radiations cancel each other to

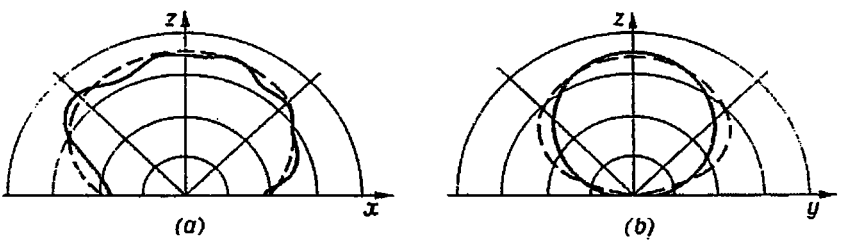


Fig. 10.17. Radiation patterns of a microstrip antenna in (*a*) the E-plane and (*b*) the H-plane: _____ experimental; — — — — — — — — calculated

a considerable degree (the exact cancellation occurs in the zOx and zOy planes). Using the technique for the calculation of antenna radiation fields formulated in Sec. 7.3, the following equations can be written for the radiation pattern of a rectangular microstrip antenna:

$$\dot{f}_\theta(\theta, \varphi) = \cos v \sin \varphi (\sin u/u)$$

$$\dot{f}_\varphi(\theta, \varphi) = \cos v \cos \varphi \cos \theta (\sin u/u) \left[1 - \frac{(\beta_0 L \sin \theta)^2}{\pi^2 - (\beta_0 L \cos \theta)^2}\right]$$

where

$$u = 0.5\ \beta_0 w \sin \theta \cos \varphi$$

$$v = 0.5\ \beta_0 L \cos \theta$$

$$\beta_0 = 2\pi/\lambda_0$$

In the equation for $\dot{f}_\varphi$, the first term in the brackets corresponds to the radiation from the two end slots with a uniform magnetic current distribution, whereas the second term represents the radiation from the lateral slots. Recall that the dimension L should be specified so as to ensure the condition of resonance, that is, such that

$$L \approx 0.5\ \lambda_0 (\varepsilon_r)^{-1/2}$$

The typical radiation pattern of a rectangular microstrip antenna is shown in Fig. 10.17. Since both end slots contribute to the overall radiation, a rectangular microstrip antenna has a directivity

somewhat exceeding that of a one-sided slot half-wave antenna, being anywhere between 3.5 and 7.0 (the lower values of the directivity correspond to the higher values of the permittivity of the substrate, which permits the radiating sheet to be made smaller in size, but impairs the directional properties).

In evaluating the efficiency of strip antennas, it is important to take into account three forms of power loss:

(1) Ohmic losses in the radiating sheet and the ground plane.

(2) Losses in the non-ideal dielectric substrate.

(3) Diversion of some input power to the arising surface wave (the dielectric substrate is in effect a slow-wave structure capable of sustaining such a wave).

Taken together, these three factors bring down the efficiency to anywhere between 50 and 80%, the lower values being associated with thinner substrates.

A major limitation of a single microstrip radiator is a limited bandwidth related to the resonant behaviour of the antenna. It has been found that, neglecting ohmic losses, the Q-factor of a square microstrip antenna, the so-called *radiation* Q, is approximately given by the following relation:

$$Q_{\Sigma} \approx \varepsilon_r D \, (\lambda_0/16h)$$

where D is the directivity, and h is the thickness of the substrate. Typically, the radiation Q for single radiators ranges between 30 and 80 in value.

The input impedance of a microstrip radiator can readily be adjusted by varying the location of the feed point. The input impedance of a rectangular antenna is a maximum when the feed point is located at the edge (being several hundred ohms in value) and varies as $\sin^2(\pi a/L)$, with the feed point located at a distance a from the centre of the antenna. The proper choice of the location for the feed point, while maintaining the condition of resonance, makes the use of any matching device unnecessary.

Microstrip antennas can readily be adapted to radiate circularly polarized waves. To achieve this with a square radiator, it is necessary to excite the cavity so that it sustains waves of two orthogonal linear polarizations with the same amplitude and with a phase shift of $\pm \pi/2$. In the arrangement of Fig. 10.18*a*, this purpose is served by a 3-dB quadrature directional coupler in which the individual and decoupled ports correspond to the radiations of waves with the polarization vector rotating clockwise and anti-clockwise. In the arrangement of Fig. 10.18*b* the feed point is positioned in a corner of the radiating sheet, and its lateral dimensions are somewhat different. In consequence, the resonance frequencies of the antenna for the two orthogonal linear polarizations lie above and below the operating frequency, respectively, the difference between the frequ-

encies being chosen so as to obtain a quadrature phase shift between the waves of the two linear polarizations. In the arrangement of Fig. 10.18c, the modes excited in the microstrip resonator by the feed line correspond to a linearly polarized radiation. Orthogonal linearly polarized waves are excited by an inclined slot cut in the metal radiating sheet of the antenna. The coupling slot 'splits' the resonance frequencies for the two linear polarizations, and if this

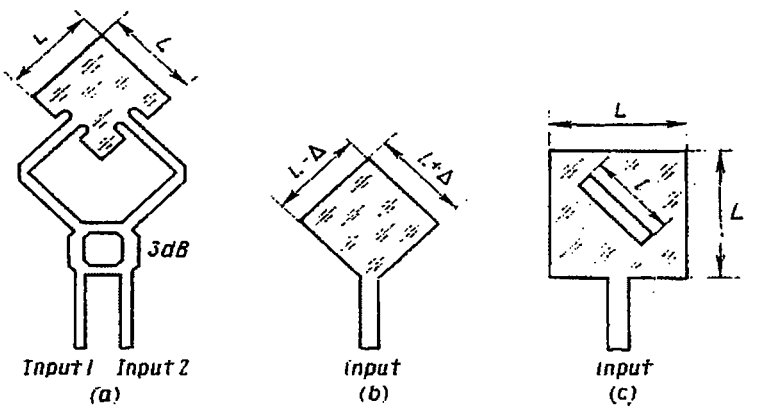


Fig. 10.18. Methods of producing circular polarization in microstrip antennas

'split' is chosen correctly, the condition can be satisfied for the radiation to be circularly polarized at the centre frequency.

At present big strides are being made in both the theory and technology of microstrip antennas which are coming in a wide variety of configurations suitable for use either as 'stand-alones' or as array elements. The frequency band in which microstrip antennas are most commonly used extends from 300 MHz to 20 GHz. Theoretical and experimental effort is mainly concentrated on developing computer-aided design techniques for microstrip antennas in combination with feeder systems, improvements in radiators and arrays, bandwidth extension, efficiency improvement, and the drive into the millimetric wavelength (EHF) band.

10.5. Principles of Frequency-Independent Antennas

The class of frequency-independent antennas comprises radiators of low directional properties, capable of operation over frequency spans with a maximum-to-minimum frequency ratio of five to one or even more, provided a high quality of input match is preserved and the shape of the radiation is not distorted too much. Such anten-

nas are admirably suited for broadband radio communication, electronic reconnaissance and electronic countermeasure systems.

The principle pivotal to understanding the operation of frequency-independent antennas is the principle of *electrodynamic similitude* or *scaling*. According to this principle, two distinct lossless antennas operating, respectively, at frequencies ω_1 and ω_2 will have the

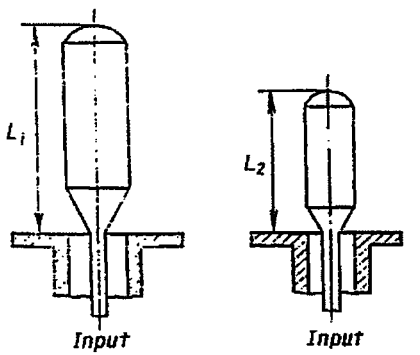


Fig. 10.19. Electrodynamically similar dipole antennas

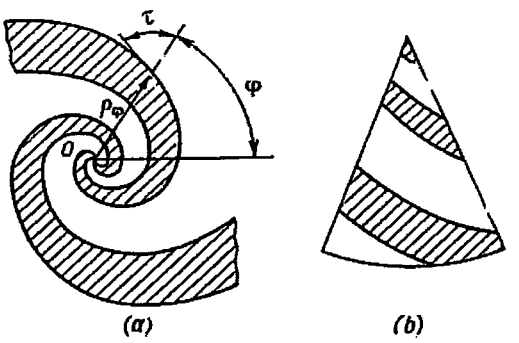


Fig. 10.20. Unbounded structures specified in angular parameters: *a*, flat; *b*, on a conical surface

same radiation pattern and the same input impedance if the antennas are the same in shape (Fig. 10.19) and the respective dimensions are inversely proportional to the ratio of frequencies

$$L_1/L_2 = \omega_2/\omega_1$$

Antennas having the same dimensions measured in wavelengths are called *electrodynamically similar*.

An only antenna that will remain electrodynamically similar to itself (or self-similar) at all frequencies can be built only if its shape can be specified entirely in angles. The simplest example is an infinitely long dipole with conically shaped arms. Other configurations specified entirely in spherical coordinates by a parametric equation

$$R = g(\theta, \varphi)$$

and solely decided by angular dimensions can be found by solving the equation

$$\mathscr{K} g(\theta, \varphi) = g(\theta', \varphi')$$

This equation presumes that the shape of the antenna, $g(\theta, \varphi)$, will, on scaling up by a factor of $\mathscr{K}$ in radius, remain the same as the original one, but turned in space. The available designs include, for example, three-dimensional structures made up of equiangular

spirals specified by the parametric equation

$$R(\theta, \varphi) = \exp[a(\varphi + \varphi_0)] T(\theta)$$

where $T(\theta)$ is an arbitrary function (a constant in a special case).

An example of an infinite plane equiangular structure formed by a pair of metal arms is shown in Fig. 10.20*a*. The edges of the arms are defined by the equation of equiangular spirals

$$\rho(\varphi) = A \exp a(\varphi + \varphi_0)$$

where A and φ_0 are constants which define the scale along the coordinate ρ and the initial angular position of the arm, whereas the parameter a is equal to the cotangent of the angle τ between the radius and a line tangent to the helix. Near the centre, the arms of the structure are cut off, and the driving voltage is fed into the gap thus formed. Note that three-dimensional structures made up of equiangular spirals can be built by placing metal arms on the surface of a cone (Fig. 10.20*b*).

A distinction of both plane and spatial equiangular spiral structures is *automatic cut-off of radiating currents*. It consists in that the currents arising in the spiral arms due to the action of the generator are abruptly limited in amplitude (by a factor of 100 or even more) as they traverse one turn of the spiral, usually equal in perimeter to one wavelength. The outer part of the structure remaining unexcited may be discarded, and this has no effect on either the radiation pattern or the input impedance of the antenna. The current cut-off was first discovered experimentally and later verified by electrodynamic analysis.

Because of current cut-off, only the central part with a diameter of $\lambda/3$ (the so-called *active region*) contributes to the radiation produced by an equiangular spiral structure of finite dimensions. As the operating frequency is changed, the electrical dimensions of the active region remain constant — a fact which makes it possible to build antennas which have a nearly constant input impedance and a negligibly varying radiation pattern over a frequency span with a maximum-to-minimum frequency ratio of 10 to 1 or even more. The lower limit of the frequency span is chosen such that the wavelength is equal to the perimeter of the last turn in the structure. The upper limit of the frequency span is set by the fact that the excitation region of the antenna, where the geometry of the spiral lines is disturbed by connection of the feed line, is comparable with the wavelength.

In the case of flat equiangular spiral antennas and in some other cases, one more factor related to the principle of permutational duality is of importance. The point is that a flat structure specified entirely in angles may at the same time be treated as an electric radiating system (a dipole) and as a magnetic radiating system

(a slot)*. Let its input impedance be $\dot{Z}_1$ and note that, as follows from the duality principle, when the metallic and slot sections of the structure are interchanged, its impedance also changes to become equal to $\dot{Z}_2$ given, according to (9.17), by the following equation:

$$\dot{Z}_2 = Z_m^2/4\dot{Z}_1 = (60\pi)^2/\dot{Z}_1$$

When the electric section and the complementary magnetic section of a structure are the same in shape, $\dot{Z}_1$ becomes equal to $\dot{Z}_2$, and the input impedance turns out to be

$$\dot{Z}_1 = \dot{Z}_2 = 60\pi, \text{ ohms}$$

at any frequency. A very simple example of how this condition is satisfied is shown in Fig. 10.21.

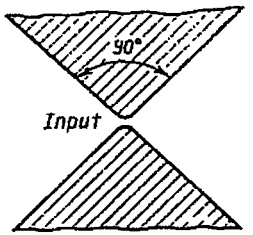


Fig. 10.21. Explaining the principle of complementary structures

Thus, the following principles applicable to frequency-independent antennas may be stated.

(1) The antenna should satisfy the condition of automatic radiating-current cut-off ensuring that the radiating part of the antenna retains an unchanged electrical dimension (the current cut-off pinciple).

(2) The antenna shape must be specified primarily by angular dimensions.

(3) The slot part of the antenna must be the same in shape as the dipole part (the principle of complementary structures).

Most important of the three is the current cut-off principle. The violation of the second and third principles in real antennas will not necessarily impair the frequency response of the antenna.

10.6. Frequency-Independent Spiral Antennas

Figure 10.22 shows the simplest spiral or flat-helix broadband antenna whose arms are bounded by four logarithmic spirals

$$\rho_1 = A \exp(a\varphi)$$
$$\rho_2 = M\rho_1$$
$$\rho_3 = A \exp[a(\varphi - \pi)]$$
$$\rho_4 = M\rho_3$$

Such an antenna cut in a metal sheet of large dimensions is energized by a flexible coaxial cable laid along one of the arms. With as many as 1.5 to 2.5 turns, the radiation pattern of the antenna

* The two make up what is known as complementary structures or antennas.—*Translator's note.*

displays two broad lobes whose maxima are normal to the plane of the antenna. The radiated field in the direction of the major lobes, as measured over a frequency span with a maximum-to-minimum frequency ratio of 20 to 1, is elliptically polarized with an axial ratio of at least 0.5. Because a change in frequency for this type of antenna is equivalent to its rotation about the axis at right angles to the plane of the sheet, the beamwidth does not remain constant — rather, it varies in a periodic fashion between the limits of 40-50° over the entire bandwidth.

The SWR of a 50-ohm coaxial line at any frequency within the bandwidth does not exceed 2.0. Variations in the parameters a and M do not markedly affect the performance of the antenna over the bandwidth and only cause a change in the mean value of the input impedance. It has been found by experiment that good results are obtained with $a = 0.30$ and $M = 0.62$.

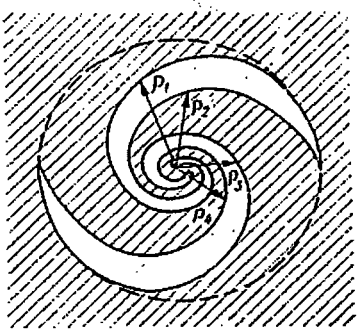


Fig. 10.22. Equiangular spiral antenna

Among the demerits of equiangular spiral antennas are the complex shape and a noticeable instability of their parameters with changes in frequency. These demerits are avoided in flat or Archimedes spiral antennas which depend for their operation on automatic current cutoff. Such antennas are simpler in shape and, what is especially important, their parameters are more stable over the bandwidth. Our discussion will be limited to the simplest antennas using two-start Archimedes spirals (or flat helices), although in practice the number of starts may be as great as four or even more.

The arms of a two-start Archimedes spiral antenna are described by the following equations:

$$\rho_1(\varphi) = a\varphi + b$$

and

$$\rho_2(\varphi) = a(\varphi - \pi) + b$$

where ρ and φ are polar coordinates, whereas a and b are constants.

This antenna is made of two conductors which may notably be flat in shape and printed on a thin sheet of an r.f. dielectric material (Fig. 10.23). The spacing between the conductors $\Delta\rho$ and the conductor width Δ remain the same within the angle φ, with the conductor spacing usually taken to be equal to the conductor width (the complementarity principle). The currents that energize the antenna can be in-phase or in anti-phase. In the case of *anti-phase excitation*,

the starts of spirals A and B are joined to a two-wire line or they can be energized by a coaxial cable laid along one of the arms. The directions of currents at the starting points of the spiral are shown in Fig. 10.24*a*. In the case of in-phase excitation, the starts of spirals, *1*, are shorted together and connected to the inner conductor of a coaxial cable *2* (Fig. 10.24*b*). The outer conductor of the coaxial cable is connected to a counterpoise in the form of a flat metal disc. The counterpoise may alternatively be printed on the reverse side of the dielectric sheet, *3*, which carries the spiral conductors. The central part of the in-phase antenna and the counterpoise form between them an unbalanced (unshielded) strip line which acts as a distributor and does not contribute to the radiation. The directions of currents at the starts of the spirals in the case of in-phase excitation are shown in Fig. 10.24*b*.

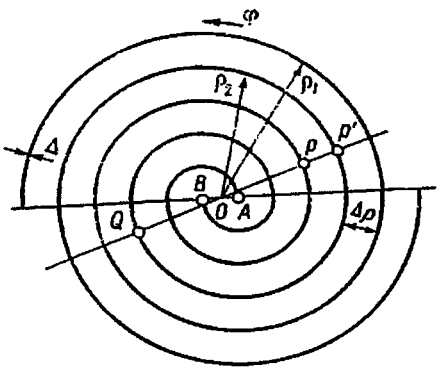


Fig. 10.23. Two-start Archimedes spiral

Let us examine the mechanism by which Archimedes spiral antennas radiate from a physical point of view. To a first approxi-

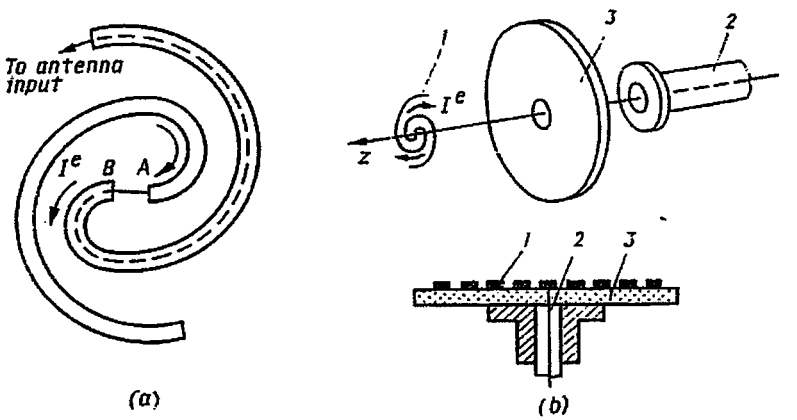


Fig. 10.24. Archimedes spiral antennas: (*a*) anti-phase and (*b*) in-phase excited

mation, the antenna may be treated as a two-wire line coiled into a spiral. Suppose that each conductor of the spiral sustains a travelling current wave and that the spacing $\Delta\rho$ between the conductors is substantially smaller than the wavelength. Let us also elucidate

the phase relationships between the currents in the spiral conductors in the case of anti-phase and in-phase excitation.

Anti-phase excitation. Let the currents at the starts of the spiral (see Fig. 10.23) be such that $\Phi_A = \pi$ and $\Phi_B = 0$. Select two points located on the different arms of the spiral, but within the same distance l of points A and B. The phase delay at these selected points relative to points A and B will be the same and equal to βl. In consequence, we may write that the phases of the currents at points P and Q are

$$\Phi_P = -\beta l + \pi$$

and

$$\Phi_Q = -\beta l$$

Let points P and Q lie on a circle of radius ρ. Consider point P' which is adjacent to point P but lies on the other conductor of the spiral. With $\Delta\rho \ll \lambda$, there will be a further phase delay at point P' relative to point Q, equal to $-\pi\beta\rho$. In consequence, at point P' the phase of the current will be

$$\Phi_{P'} = -\beta l - \pi\beta\rho$$

The difference in phase, $\Delta\Phi$, between the currents in the adjacent conductors of the spiral, as measured at points P' and P, will be

$$\Delta\Phi = \Phi_P - \Phi_{P'} = \pi\beta\rho + \pi$$

Let us find the value of the radius ρ_n for which

$$\Delta\Phi = 2\pi n, \quad n = 1, 2, \ldots$$

and the currents in the adjacent conductors are in phase at points P and P'. The applicable condition has the form

$$2\pi\rho_n = (2n - 1)\lambda, \quad n = 1, 2, \ldots$$

As follows from the above condition, the currents in the adjacent conductors of the spiral, as measured on circles with a perimeter equal to an odd number of wavelengths, are in phase and can produce a strong electromagnetic radiation. The radiation intensity is at its highest from the region occupied by the circle whose radius is $\lambda/2\pi$, but on passing through the successive turns with such a radius the currents in the conductors of the spiral line decrease in magnitude abruptly. The subsequent turns of the spiral for which the in-phase condition is satisfied at $n = 2, 3, \ldots$ remain practically unenergized and do not affect the performance of the antenna. Thus, the active region of an Archimedes spiral antenna which is energized in anti-phase may arbitrarily be replaced by a single turn of radius $\lambda/2\pi$ (a one-wavelength loop) in which the equivalent electric

current distribution obeys the travelling-wave law

$$\dot{I}^{e}_{\varphi,\,ap}(\varphi) = \dot{I}_0 \exp(-j\varphi) \tag{10.1}$$

Here the subscript "ap" stands for "anti-phase". The perimeter of this equivalent turn is equal to one wavelength.

Since the reasoning set forth above holds at any wavelength, an anti-phase spiral antenna is frequency-independent in terms of both input impedance and radiation pattern, provided the operating frequency is in excess of some limiting value which depends on the outside diameter of the antenna.

In-phase excitation. Applying a similar reasoning to the phase relationships between the currents in the conductors of an in-phase excited spiral antenna, it is an easy matter to see that the currents in the adjacent conductors are in phase on circles of radius ρ_m, such that

$$2\pi\rho_m = 2m\lambda, \quad m = 1, 2, \ldots$$

Because of the high radiation intensity in the region taken up by the circle of radius λ/π, automatic current cut-off occurs, and the turns of radii greater than λ/π remain unexcited. Thus, the active region of an Archimedes spiral antenna may be replaced by a single turn of radius λ/π (a two-wavelength loop) in which the equivalent current distribution obeys the travelling-wave law

$$\dot{I}^{e}_{\varphi,\,ip}(\varphi) = I_0 \exp(-j2\varphi) \tag{10.2}$$

Here the subscript "ip" stands "in-phase". The operating range of this antenna is defined such that

$$R_{cp} < \lambda/\pi < R_{out}$$

where R_{out} is the outside radius of the antenna, and R_{cp} is the radius of the metallic counterpoise in the feed.

Radiation patterns of Archimedes spiral antennas. In the calculation of the radiation pattern, it is convenient to place an Archimedes spiral antenna in the equatorial plane of a spherical coordinate system (such that its origin lies at the centre of the spiral). Then, on assuming that the antenna owes its radiation to extraneous currents as defined by Eq. (10.1) or (10.2), and invoking the field calculation technique set forth in Sec. 7.3, the following equations can be written for the radiation pattern:

$$\dot{f}_\theta(\theta, \varphi) = j\,[J_{n-1}(n \sin\theta) + J_{n+1}(n \sin\theta)]\cos\theta \exp(-jn\varphi) \tag{10.3}$$

$$\dot{f}_\varphi(\theta, \varphi) = [J_{n-1}(n \sin\theta) - J_{n+1}(n \sin\theta)] \exp(-jn\varphi)$$

where $J_{n\pm1}(x)$ is a Bessel function, with $n=1$ applying when the spiral is excited in anti-phase, and $n = 2$ applying when the spiral is excited in-phase.

Equations (10.3) also provide a tool with which to calculate the radiation of circular loop antennas with a radius equal to $a = (\lambda/2\pi)n$, and with the extraneous current distributed as $\dot{I}^i_\varphi \equiv \exp(-jn\varphi)$, where $n = 1, 2, \ldots$.

The radiation patterns lying in the meridional plane $\varphi = 0$, as calculated by Eqs. (10.3), appear in Fig. 10.25. In view of the symmetry of the antenna, the radiation patterns are plotted for only one quadrant. Complete three-dimensional radiation patterns are bodies of revolution about the z-axis of the antenna. As follows from Eqs. (10.3), the components $\dot{f}_\theta$ and $\dot{f}_\varphi$ are in phase quadrature everywhere. Therefore, the total radiated field of the antenna has predominantly an elliptical state of polarization. In the plane of the spiral, however, the component $\dot{f}_\theta$ always vanishes, so the resultant field is linearly polarized. A purely circular state of polarization arises in the axial direction $\theta = 0$ (or $\theta = \pi$) in the case of anti-phase excitation. With in-phase excitation, there is no radiation along the axis.

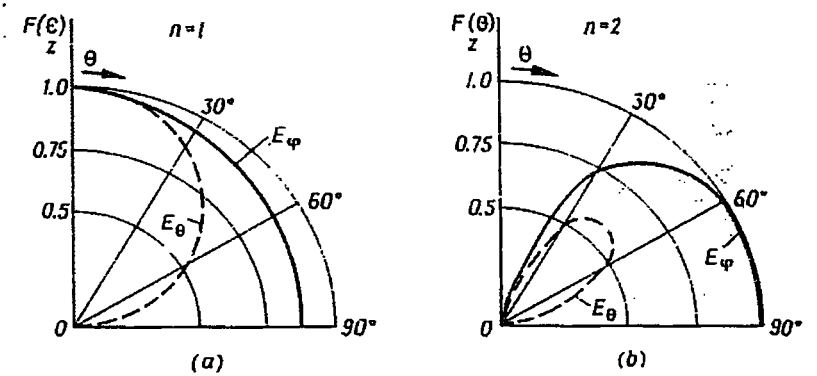


Fig. 10.25. Radiation patterns of Archimedes spiral antennas: *a*, anti-phase excitation; *b*, in-phase excitation

Archimedes spiral antennas could be adapted to produce a one-sided radiation. This could be done by placing a screen or a closed box on the underside of the spiral. This would, however, make the antenna parameters frequency-dependent to a greater extent and reduce its bandwidth.

10.7. Log-Periodic Antennas

Automatic cutoff of radiating currents is also embodied in antennas whose parameters vary periodically with the logarithm of the frequency, giving rise to the description 'log-periodic'. One of the early log-periodic antennas is shown in Fig. 10.26. Its arms are structures of alternating dipoles in the form of ridges and recesses. The geometry of these structures is specified by a dimensionless period, or geometric ratio

$$\tau = R_n/R_{n+1}$$

by the shape factor

$$\sigma = r_n/R_n$$

and by the angles α and β. Most often, use is made of structures for which the shape factor is equal to the square root of the geometric ratio, $\sigma = \sqrt{\tau}$. At the origin of coordinates, the antenna arms are separated by a gap. The braid of the coaxial feeder cable is soldered to one arm of the antenna all along its length, and the central conductor is connected to the start of the other arm.

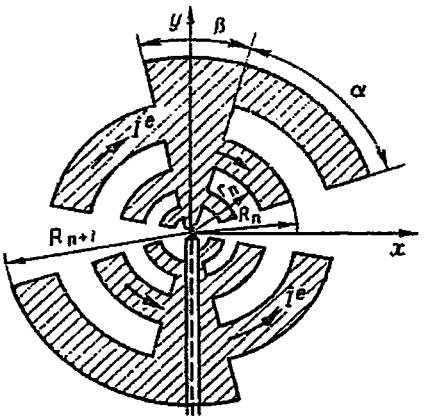


Fig. 10.26. Flat log-periodic antenna

As has been found by experiment, the electric current in the ridges of the structure increases on moving away from the centre and reaches a maximum at the location of resonant ridges of about a quarter-wave long. The directions of the strongest currents which contribute the most power to the radiated field are shown by the arrows in Fig. 10.26. On passing beyond the resonant ridges, the currents decrease in magnitude abruptly — this is what we have called 'current cut-off'. As the radiated field goes down in frequency, resonance occurs at progressively more distant ridges. The ratio of the two frequencies at which the adjacent ridges resonate is exactly equal to the ratio

$$\tau = \omega_1/\omega_2$$

Because of this, the input impedance and radiation pattern of the antenna are said to be periodic with the logarithm of the frequency.

With a proper choice of the structure parameters, the performance of the antenna during one period will change only slightly. The exact replication of this change from one period to the next ensures a satisfactory behaviour of the antenna characteristics over a very broad bandwidth. To a first approximation, the operating bandwidth is bounded by the resonance frequencies of the outer ridges closest

to the centre of the antenna, on the one hand, and those most remote, on the other. It is quite an easy matter to obtain bandwidths up to 10 : 1 and even higher.

The average input impedance of a log-periodic antenna does not usually exceed 100 ohms and depends on the angles α and β. With $\alpha = \beta = 45°$, the electric-dipole part of the antenna is the same in shape as its complementary slot part. Therefore, in accordance with the equation

$$\dot{Z}_1 \dot{Z}_2 = (60\pi)^2$$

the input impedance must be around 180 ohms. Measurements yield a somewhat lower figure equal to around 150 ohms. The average

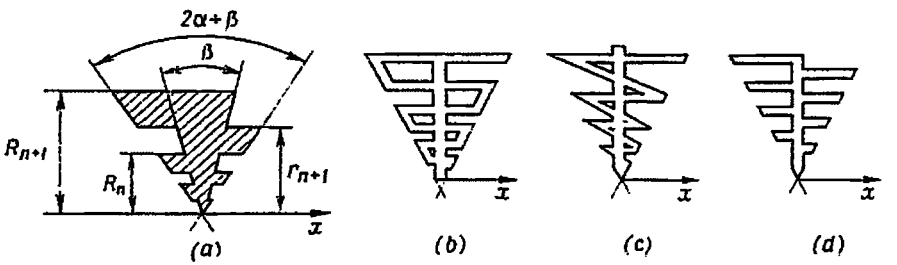


Fig. 10.27. Log-periodic structures

input impedance can be reduced so as to facilitate the proper match between the antenna input and a 75-ohm coaxial cable by reducing the angle β.

The radiation pattern of a flat log-periodic antenna has two major lobes which are at right angles to the plane of the arms. In the plane of the antenna, the radiation is negligible in all directions. The E vector in the direction of the major lobe is polarized parallel to the x-axis, whereas the radiation with a state of polarization parallel to the y-axis is low. The width of the lobes in the radiation pattern depends on the dimensionless period (geometric ratio) of the structure, τ. At $\alpha = \beta = 45°$, the decrease in τ from 0.80 to 0.25 brings on a gradual decrease in the half-power lobe width from 73° to 38°.

Advances in the design of log-periodic antennas have led to simpler arm shapes (only a half of an antenna is shown in Fig. 10.27*a* through *d*). It has also been found that placement of the antenna arms so that the included angle ψ relative to the x-axis is less than 90° can assure a unidirectional radiation towards the apex of the structure (Fig. 10.28*a*). The direction of the major lobe is the same as that of the bisector of the angle ψ.

In the limiting case of $\psi = 0$, the two arms of the antenna merge together to form a flat log-periodic antenna (Fig. 10.28*b*). It is convenient to treat such an antenna as a linear array of symmetrical

dipoles of a monotonically decreasing length, energized by a two-wire line with crossed conductors. The line is energized by a ge-

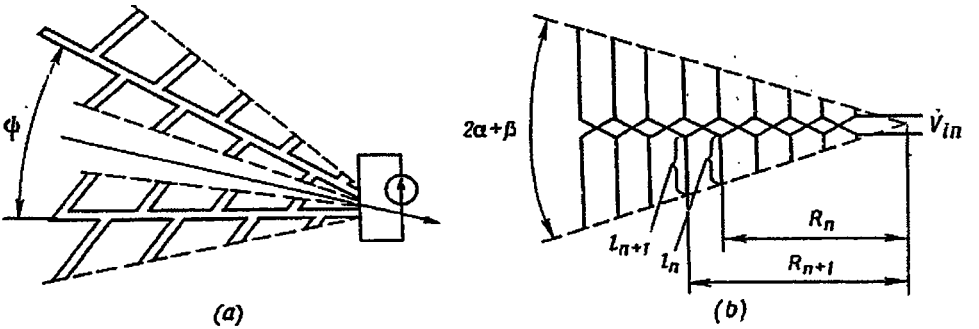


Fig. 10.28. Unidirectional log-periodic dipole antennas: *a*, spatial; *b*, planar

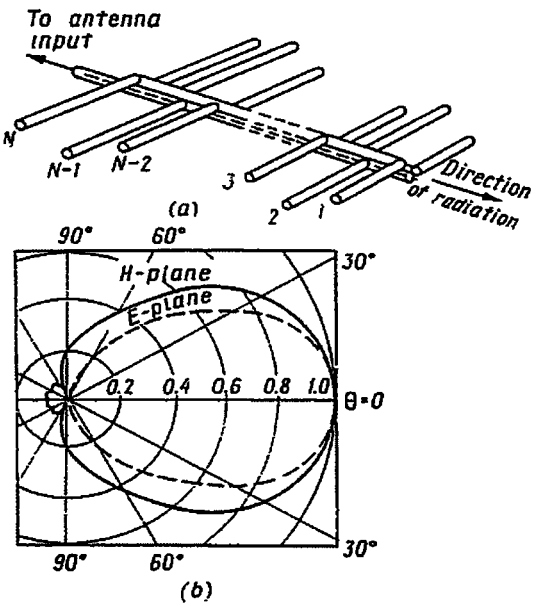


Fig. 10.29. Log-periodic dipole antenna

nerator located on the side of the shorter dipoles. The arrangement of a flat antenna which needs no balancing device is shown in Fig. 10.29*a*.

For insight into operation of a log-periodic dipole antenna, it is necessary, first of all, to isolate the active portion containing the strongest radiating currents. To a first approximation, this portion

includes the resonant-length dipole (around a quarter-wavelength long) and the two dipoles adjacent to it from both sides. All the other dipoles are well off resonance, they are energized very little, and their contribution to the radiated field is small.

The conditions of excitation and the cross coupling between the dipoles within the active portion are combined in such a way that the longest dipole behaves as a reflector and the current it carries leads in phase the current carried by the resonant dipole. In contrast, the current in the shortest dipole lags in phase behind the current in the resonant dipole, and the shortest dipole acts as a director. The fields radiated by the several dipoles in the active portion combine together in the direction of the antenna apex and subtract from one another in the reverse direction. The shape of the measured radiation-pattern in the **E** and **H** planes of a flat antenna is shown in Fig. 10.29*b*. In the **E** plane, the radiation pattern is narrower due to the directional behaviour of one dipole in this plane.

The lower limit for the bandwidth of a log-periodic dipole antenna is set by the maximum permissible dimensions of the largest dipoles ($l_{max} \approx \lambda_{max}/4$), and the upper limit by the dimensions of the smallest dipole ($l_{min} \approx \lambda_{min}/4$). However, in order to preserve good directional properties at the extreme frequencies of the bandwidth, care must be taken to provide one or two more longer reflector dipoles past the dipole which resonates at the longest wavelength, and two or three more director dipoles ahead of the dipole which resonates at the shortest wavelength. In practice it is possible to build dipole log-periodic antennas which preserve nearly constant directional properties with an input reflection coefficient of $|\dot{\rho}| \leqslant 0.20$ to 0.25 over a 10 : 1 bandwidth.

The portion of the two-wire line between the driving point and the start of the active portion is used to channel power. The short dipoles located within these portions have a high reactance and load the line little, only slightly decreasing its equivalent characteristic impedance. The power carried by the two-wire feed line is intensively radiated by the active portion, which is why the segment of the line extending from the apex of the antenna to the start of the active portion operates in the travelling-wave mode.

Chapter Eleven

Linear Antennas and Arrays

11.1. Linear Radiating Systems. The Ideal Linear Radiator

The term 'linear radiating system' refers to an assemblage of identical electromagnetic-field sources, called *elements*, arranged continuously or discretely in a specified direction in space. Exam-

ples are a straight current-carrying wire, an extended slot in a plane screen with a voltage applied between its edges, an array of identical dipoles with their centres lying on a straight line, to name but a few. The longitudinal dimension of a linear radiating system is arbitrary, whereas the transverse dimensions are most often small or comparable with the wavelength. The individual radiation patterns of the constituent elements of a system are assumed to be the same, which is equivalent to postulating the same radiating current distribution in all the elements. On the whole, a linear radiating system is fully specified by the distribution of the radiator centres along the axis and by the distribution of the complex amplitudes of excitation currents among the individual elements, or what we will call the excitation or current distribution in amplitude and phase.

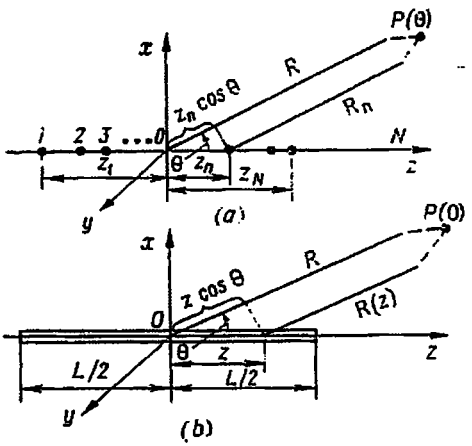


Fig. 11.1. To calculation of the array factor

As follows from the product pattern (or pattern multiplication) theorem formulated in Sec. 9.6, the radiation pattern of a linear radiating system may be written as

$$\dot{\mathbf{f}}(\theta, \varphi) = \dot{\mathbf{F}}_{el}(\theta, \varphi)\, \dot{f}_{\Sigma}(\theta)$$

where $\dot{\mathbf{F}}_{el}(\theta, \varphi)$ = complex vector radiation pattern of one single element in its own coordinate system, or the element factor

$\dot{f}_{\Sigma}(\theta)$ = complex scalar pattern factor of an array of isotropic point sources located at the positions of the elements along the axis of the system or, simply, the array factor

For a system of discrete radiators located at N specified points z_n on the axis, that is, for a linear array (Fig. 11.1*a*), the array factor $\dot{f}_{\Sigma}(\theta)$ may, in accordance with Eq. (9.22), be written as

$$\dot{f}_{\Sigma}(\theta) = \sum_{n=1}^{N} \dot{I}_n \exp(j\beta z_n \cos\theta) \tag{11.1}$$

where

$$\dot{I}_n = I_n \exp(j\Phi_n)$$

is the complex amplitude of the excitation current applied to the nth radiator, and $z_n \cos\theta$ is the path length difference between the rays that pass from the origin of the common coordinate system and the position of the nth radiator to the observation point.

A continuous linear radiating system, or a *linear radiator*, may be treated as the limiting case of a linear array of length L and with a number of elements tending to infinity (Fig. 11.1*b*). Then the summation in (11.1) can be replaced by integration, and the array factor of the linear radiator takes the form

$$\dot{f}(\theta) = \int_{-L/2}^{L/2} \dot{I}(z) \exp(j\beta z \cos\theta)\, dz \qquad (11.2)$$

where

$$I(z) = I(z) \exp[j\Phi(z)]$$

is the current distribution in amplitude and phase along the radiator and $z \cos\theta$ is the path length difference.

The array factors of discrete and continuous radiating systems do not depend on the azimuthal coordinate and have the property of rotational symmetry about the z-axis. Physically, these factors describe the *interference of the spherical waves* excited by isotropic point sources. In designing pencil-beam antennas, it is usually sought to arrange the elements so that their radiated fields combine in phase or nearly in phase to produce a maximum of radiation in the specified direction and cancel one another as much as possible in all other ("null") directions. If the radiation pattern of one element is sufficiently broad, it will not affect materially the shape of the overall radiation pattern of the antenna in the vicinity of the radiation maximum. Therefore, in analysis of pencil-beam antennas, main emphasis is placed on the array factor.

On introducing a new variable

$$\varkappa = \beta \cos\theta$$

called the *spatial frequency*, and formally extending the limits of integration to infinity (for which purpose it will suffice to complete the definition of the current distribution function with zero outside the interval L), Eq. (11.2) may be re-cast as an integral Fourier transform

$$\dot{f}(\varkappa) = \int_{-\infty}^{\infty} \dot{I}(z) \exp(j\varkappa z)\, dz \qquad (11.3)$$

Thus, the array factor and the excitation function of a linear radiator are connected in the same manner as the shape of a voltage pulse in an electric circuit and its frequency spectrum. This analogy extends far enough for many findings in the spectral theory of sig-

nals to be carried over into antenna theory. Another advantage stemming from writing the array factor of a linear antenna in the form of Eq. (11.3) is the fact that calculations can be carried out on a computer, no matter how complex the amplitude-and-phase excitation distributions may be, using standard Fast Fourier Transform (FFT) programs which speed up computations tens or even hundreds of times as compared with programs for conventional numerical integration.

The ideal linear radiator. This term refers to a linear radiating system for which the current distribution has the form

$$\dot{I}(z) = I_0 \exp(-j\xi\beta z) \text{ with } |z| \leqslant L/2 \qquad (11.4)$$

where I_0 = constant amplitude
$\beta = 2\pi/\lambda$ = wave number of the medium around the radiator
$\xi = c/v$ = relative phase velocity, or the ratio of the velocity of light, c, to the phase velocity v of the excitation

An ideal radiator has a uniform current distribution and a linear phase progression typical of a wave travelling in the z-direction (Fig. 11.2). At $\xi=0$, the excitation has a constant phase along the radiator, and this corresponds to the infinite velocity of propagation for the excitation wave. At $|\xi|<1$, the phase velocity of the excitation exceeds the velocity of light in the medium surrounding the antenna. At $|\xi| = 1$, the phase velocity is exactly equal to that of light. Finally, at $|\xi|>1$ the ideal linear radiator is excited by a slowed travelling wave. The positive values of ξ correspond to the travel of the excitation wave in the positive z-direction; the negative values of ξ correspond to the travel of the excitation wave in the negative z-direction.

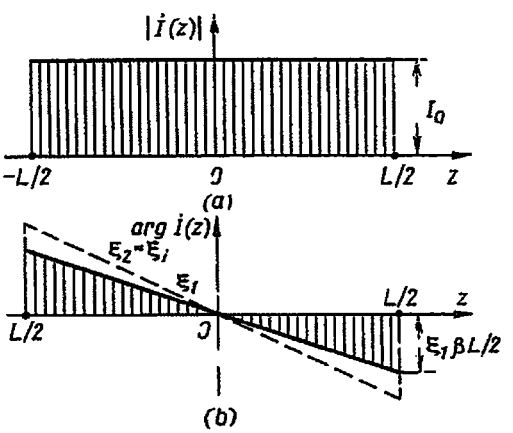


Fig. 11.2. Current distribution in an ideal linear radiator: *a*, uniform (constant) amplitude and *b*, linear phase

An ideal linear radiator is a kind of standard relative to which one can assess the properties and parameters of linear radiators with other excitation distributions. Let us find the array factor for an ideal linear radiator. On substituting (11.4) into (11.2), taking the integral, and discarding the immaterial amplitude factor I_0L, we

get

$$f(\theta) = \sin \Psi/\Psi \tag{11.5}$$

where

$$\Psi = 0.5\,\beta L\,(\cos\theta - \xi)$$

The new generalized angle variable Ψ has the significance of a half of the difference in phase between the waves arriving at a distant observation point from the extreme points of the radiator and takes into account both the spatial path length difference $\beta L \cos\theta$ and the total difference in phase between the excitations at the extreme points of the radiator.

The array factor defined by (11.5) is a real function. The equiphase surfaces of the far-zone or radiation field are spheres with centres at the mid-point of the radiator. In consequence, the phase centre of an ideal linear radiator lies at its mid-point, irrespective of the relative phase velocity of the excitation wave.

11.2. Analysis of the Array Factor of an Ideal Linear Radiator

Let us refer to Fig. 11.3 showing, at its top, a plot of the magnitude of $\sin\Psi/\Psi$ as a function of the generalized angle variable. The plot displays one *main lobe* of unity value at $\Psi = 0$, and several *side lobes* located symmetrically on both sides of the main lobe. The main lobe is produced by the in-phase addition (constructive interference) of the waves arriving from the individual elements of the system. In directions such that $\Psi \neq 0$, the waves arriving from the various elements differ in phase and interfere destructively, so that the resultant field is reduced in intensity, this reduction being greater for greater values of $|\Psi|$. The zero-power width of the main lobe (that is, one measured between reference lines of zero radiation intensity) is equal to 2π, and that of each side lobe is half as great. Radiation nulls are located at points $n\pi$, where n is equal to ± 1, ± 2, etc. The positions of the side lobes approximately correspond to the mid-points of the spacings between adjacent nulls. The sidelobe level is given by a simple formula, $1/[\pi\,(n + 0.5)]$, where n is the index of the side lobe. The level of the first side lobe (which is the largest of all side lobes) is around 0.21, or —13.2 dB.

How one passes from the generalized angle variable Ψ to the physical observation angle θ is illustrated in the plot of $\Psi(\theta)$ placed at the bottom of Fig. 11.3. On entering this plot with any specified value of θ, it is an easy matter, by simple constructions, to find the respective value of Ψ:

$$\Psi = 0.5\,\beta L\,(\cos\theta - \xi)$$

Next, one turns to the plot of $|\sin \Psi/\Psi|$ at the top and finds the array factor.

The maximum and minimum values of the function $\Psi(\theta)$,

$$\Psi_{max} = 0.5\,\beta L\,(1 - \xi)$$

and

$$\Psi_{min} = -0.5\,\beta L\,(1 + \xi)$$

limit the operating interval of the function $f(\Psi)$ which has a direct bearing on the form of the array factor. The total extent of this interval, $\Psi_{max} - \Psi_{min} = \beta L$, is solely decided by the electrical

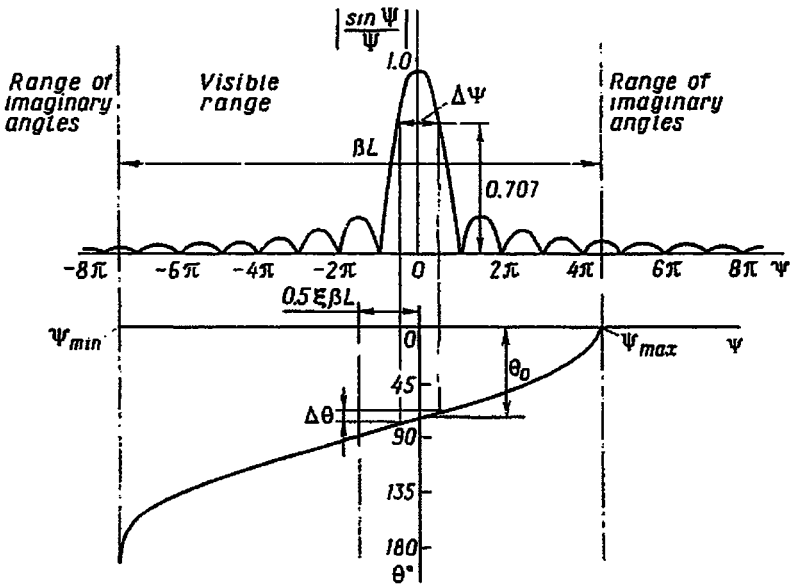


Fig. 11.3. To analysis of the array factor

length of the antenna. As the antenna length increases, a progressively greater number of lobes falls within the operating interval. The position of the mid-point of the operating interval, $-\xi\beta L/2$, is decided by the relative phase velocity. It is customary to refer to the operating interval of the function $f(\Psi)$ as the *range of real observation angles*. The values of Ψ outside the operating range correspond to $|\cos\theta| > 1$ which may be interpreted as the cosines of 'imaginary' angles. The interval of real observation angles is usually called the *visible range*. The spatial frequencies associated with it are such that $-\beta \leqslant \varkappa \leqslant \beta$.

Radiation modes of a linear antenna. The angular position of the main lobe of the radiation pattern is defined to be such that $\Psi_0 = 0$.

Hence it follows that

$$\cos \theta_0 = \xi \quad \text{or} \quad \theta_0 = \text{arc cos } \xi \quad \text{for} \quad |\xi| \leqslant 1 \tag{11.6}$$

With in-phase excitation, the angular position of the major maximum is perpendicular to the axis of the antenna, in this case the antenna is called a *broadside array* and it produces a *broadside beam.*

With $0 < |\xi| < 1$, the main lobe of the array factor deviates from the perpendicular to the axis of the antenna in the direction of propagation of the excitation wave, and a tilted beam is produced. The change in the spatial position of the main lobe is called *scanning.* As ξ varies between the limits $[-1, 1]$, the main lobe scans an angle lying in the interval $180° > \theta_0 > 0$.

With $|\xi|$ tending to unity, the main lobe starts moving beyond the visible range so that at $|\xi| = 1$ the main maximum is oriented exactly along the axis of the antenna. This is the *end-fire mode of radiation*, and the antenna is called an *end-fire array.*

With $|\xi|$ greater than unity, the main lobe occurs outside the visible range. Here, with the values of $|\xi|$ close to unity, the end-fire mode prevails as long as the part of the main lobe remaining within the visible range exceeds the first side lobe in level. As $|\xi|$ keeps increasing, only the side lobes of the function $|\sin \Psi/\Psi|$ remain in the visible range, and there is not a single direction in space where the radiations from all the elements would be in phase. The waves radiated by the individual elements of the antenna cancel one another to a considerable degree, and the antenna turns out to be a poorly radiating one.

The beamwidth of an ideal linear antenna. Because the generalized angle variable Ψ is a nonlinear function of θ, the half-power beamwidth does not remain constant as the beam scans space. Rather, it changes abruptly on passing to the end-fire mode. Let us refer to Fig. 11.3 again. As is seen from the plot at the top, the half-power beamwidth is 2.78 rad. On passing to the angle variable θ, it is essential to take into account the slope of the function $\Psi(\theta)$:

$$\Delta\theta = \Delta\Psi/(|\partial\Psi/\partial\theta|_{\theta=\theta_0}) = 2.78/0.5\beta L \sin\theta_0$$
$$= 0.886\lambda/L \sin\theta_0 = 51°\lambda/L \sin\theta_0 \tag{11.7}$$

Thus, the beamwidth decreases as the antenna length L/λ increases and as the direction of radiation moves closer to the equatorial plane, $\theta = \pi/2$. The beamwidth is a minimum in the broadside mode of radiation.

The zero-power beamwidth can be obtained in a similar manner

$$\Delta\theta_0 = 2\lambda/L \sin\theta_0 = 114° \lambda/L \sin\theta_0$$

The estimates given above are based on the assumption that the function $\Psi(\theta)$ is straightened in the vicinity of the radiation maxi-

mum. Therefore, their accuracy improves at the higher values of L/λ and for directions of radiation which are not too close to the axis of the antenna. With L greater than or equal to five wavelengths, Eq. (11.7) will be less than 0.2% in error in the case of a broadside array. In the case of a tilted (or scanned) beam, the error can rise to 4% when the beam moves towards the antenna axis to within twice its width.

Now let us consider the beamwidth in the end-fire case. To begin with, consider the case when $|\xi| = 1$, that is, when the excitation

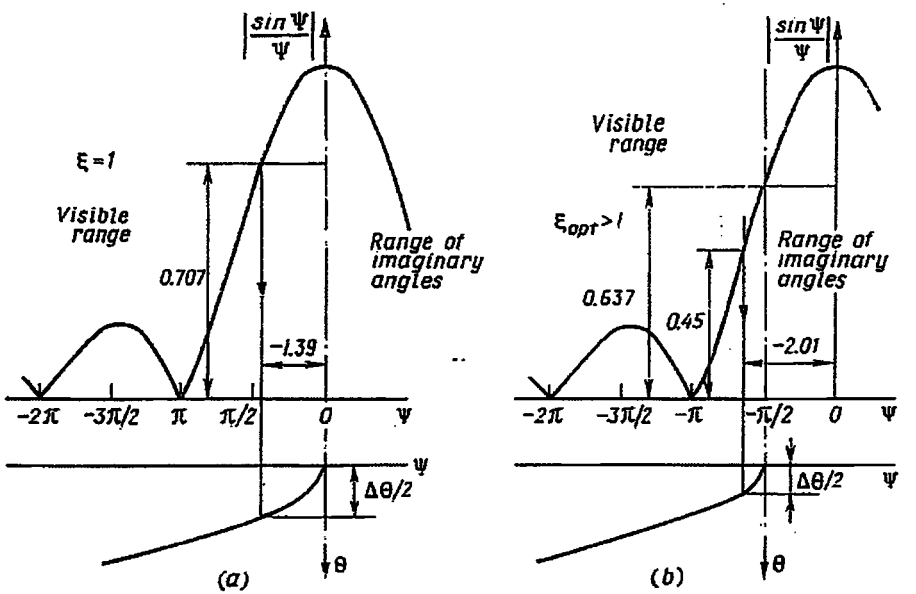


Fig. 11.4. To evaluation of the beamwidth of a linear antenna in the axial radiation mode (an end-fire antenna)

wave is propagated along the antenna at exactly the velocity of ligth. In the circumstances, the main lobe of the radiation pattern is formed as illustrated in Fig. 11.4*a*. The half-power beam width of the main lobe in terms of the variable Ψ is around 1.39 and in order to find the angular beamwidth we should solve the following equation:

$$-1.39 = (\pi L/\lambda)(\cos \Delta\theta/2 - 1)$$

At high values of the ratio L/λ, the argument of the cosine is zero very nearly. Because of this, we may well use the approximate expression

$$\cos \alpha \approx 1 - \alpha^2/2 + \ldots$$

On substituting it in the above equation, we obtain

$$\Delta\theta \approx 2\ (0.886\ \lambda/L)^{1/2} = 108^\circ\ (\lambda/L)^{1/2} \quad \text{for } |\ \xi\ | = 1 \tag{11.8}$$

which is substantially greater than is the broadside case.

In the end-fire (or axial-radiation) mode, the main lobe can be markedly reduced in width by operation at values of the relative phase velocity such that $|\ \xi\ | > 1$. Now the central part of the main lobe of the function $|\sin \Psi/\Psi|$ moves into the range of imaginary angles (Fig. 11.4*b*), and the main lobe of the array factor is sharpened appreciably. At the same time, however, there is an increase in the sidelobe level which is now determined relative to the value of $|\sin \Psi/\Psi|$ at the boundary of the visible range rather than relative to its main maximum. A satisfactory balance between the narrowing of the main lobe and the growth of the side lobes can be struck by placing the boundary of the visible range at point $\Psi\ (0) \approx -\pi/2$. As will be shown later, this condition also ensures a maximum directivity for a linear antenna with a slowed phase velocity. Therefore, we may call this the *criterion of optimality for a linear antenna with a slowed phase velocity of the excitation wave*. More commonly, it is known as the *Hansen-Woodyard condition for increased directivity**. When written out in full, it has the form

$$-\pi/2 = \pi L\ (1 - |\ \xi\ |)/\lambda$$

Hence, we may write the following relations:

$$|\ \xi_{opt}\ | \approx 1 + \lambda/2L$$

and (11.9)

$$L_{opt} \approx \lambda/2\ (|\ \xi\ | - 1)$$

Using them, one can readily determine an optimal relative phase velocity for a specified antenna length or compute an optimal antenna length for a specified relative phase velocity.

The array (or pattern) factor at the boundary of the visible range at $\Psi_{max}\ (0) = -\pi/2$ is $2/\pi$, and the first sidelobe level is one-third relative to that value, or 9.54 dB down. At a half-power point, the function $|\sin \Psi/\Psi|$ is around 0.45 which corresponds to an argument value of -2.01 rad. Hence, we may write the following equation for the half-power beamwidth:

$$-2.01 = (\pi L/\lambda)\ (\cos \Delta\theta_{opt}/2 - \xi_{opt})$$

On substituting for ξ_{opt} from Eq. (11.9) and on replacing $\cos \alpha \approx \approx 1 - \alpha^2/2$ . . ., we obtain the following design equation:

$$\Delta\theta_{opt} \approx 2\ (0.28\ \lambda/L)^{1/2} = 60.6^\circ\ (\lambda/L)^{1/2} \tag{11.10}$$

---

* It is alternatively known as the *Hansen-Woodyard increased directivity criterion.—Translator's note.*

Thus, the replacement of $|\xi| = 1$ by an optimal relative phase velocity reduces the width of the main lobe in the end-fire radiation by a factor of about 1.8.

Note that an end-fire linear radiator has a broader radiation pattern than a broadside array. For example, with L equal to ten wavelengths, the beamwidth is 5.1° with ξ equal to zero and 34° with ξ equal to unity, whereas the optimal value is $\Delta\theta_{opt} = 19°$ at $\xi_{opt} = 1.05$. However, end-fire linear antennas provide directional properties in two mutually perpendicular planes whereas in-phase linear broadside arrays concentrate the radiated power to a narrow beam solely in the equatorial plane.

11.3. The Directivity of an Ideal Linear Radiator

In order to obtain the exact value of the directivity for an ideal linear radiator, it is necessary to substitute into the general directivity equation (7.18) the complete normalized radiation pattern of the antenna, taking into account the field-strength radiation pattern of one element. As a way of simplifying computations, it is legitimate to assume that the elements are non-directional (isotropic) and to calculate the directivity solely for the array factor. Since the array factor of a linear radiator is independent of the azimuthal angle, the integral with respect to φ in the denominator of Eq. (7.18) is equal to 2π, and this simplifies the equation for the directivity as

$$D = 2f^2(\theta_0) \Big/ \int_0^{\pi} f^2(\theta) \sin\theta \, d\theta$$

where $f(\theta)$ is defined by Eq. (11.5), and $f^2(\theta_0) = 1$ for $|\xi| \leqslant 1$, where the additional factor

$$f^2(\theta_0) = \sin^2\Psi(0)/\Psi^2(0)$$

for $|\xi| > 1$ is invoked for conversion to a radiation pattern normalized to unity.

The integral in the denominator can be reduced to the form

$$\int_0^{\pi} f^2(\theta) \sin\theta \, d\theta = \frac{2}{\beta L} \int_{\Psi_{min}}^{\Psi_{max}} \frac{\sin^2\Psi}{\Psi^2} \, d\Psi$$

where the limits of integration coincide with the boundaries of the visible range (see Fig. 11.3). On re-arranging [1], the final expression for the directivity of the array factor of an ideal linear radiator takes the form

$$D = \beta L f^2(\theta_0) \Big/ \left[\frac{\sin^2\Psi_{min}}{\Psi_{min}} - \frac{\sin^2\Psi_{max}}{\Psi_{max}} + \mathrm{Si}(2\Psi_{max}) - \mathrm{Si}(2\Psi_{min}) \right] \tag{11.11}$$

where

$$\mathrm{Si}\,(x) = \int_0^x (\sin \tau/\tau)\, d\tau$$

is the integral sine.

Let us take a closer look at how the directivity depends on the relative phase velocity ξ. In the case of a broadside or tilted (scanned) beam, with $|\xi| < 1 - \lambda/L$, the main lobe of the array factor is fully located in the visible range and bounded by the values $\Psi = \pm \pi$. On substituting these values in (11.11), we obtain, in accordance with the definition in (7.19), *the directivity of a linear radiator in terms of the main lobe of the radiation pattern* as

$$D' = \beta L/2\mathrm{Si}\,(2\pi) \approx 2.22\, L/\lambda \qquad (11.12)$$

where use is made of the tabulated values of $\mathrm{Si}\,(2\pi) = -\mathrm{Si}\,(-2\pi)$, approximately equal to 1.418.

Equation (11.12) gives the upper bound on the directivity of an ideal linear antenna in the tilted-beam case, because the side lobes are neglected. The lower bound is obtained on the assumption that the antenna length tends to infinity and that the visible range encompasses an infinite number of side lobes. Then,

$$D = 2L/\lambda \quad \text{for } L \gg \lambda \qquad (11.13)$$

The overall directivity of an antenna and its directivity in terms of the main lobe are connected by an equation of the form

$$D = D'\,(1 - \beta_s)$$

where β_s takes care of the power scattered by the side lobes. On comparing (11.12) and (11.13), we may conclude that the power scatter factor for an ideal linear radiator does not exceed 10%, tending to this value as the ratio L/λ tends to infinity.

Thus, the quantity $2L/\lambda$ at $L \gg \lambda$ is the standard value of directivity for an ideal linear antenna in the broadside and tilted-beam radiation modes and is *independent of the scanning direction.* At first glance this might seem strange because the deviation of the beam from the normal to the antenna usually causes the width of the main lobe to vary as $1/\sin \theta_0$, and broadening of the main lobes ordinarily impairs directivity. The explanation lies in the fact that the deviation of the main lobe from the equatorial plane of the antenna is accompanied by a decrease in the solid angle per meridional degree,

$$d\Omega = \sin \theta_0\, d\theta\varphi$$

and this balances out the tendency for the beam to broaden as $1/\sin \theta_0$.

As the relative phase velocity increases from $1 - \lambda/L$ to unity, the main lobe gradually moves towards the axis of the antenna and the directivity increases. At $\xi = 1$ we have $\Psi_{max} = 0$, so that the upper and lower boundaries of the directivity turn out to be respectively as follows:

$$D' = 4.44\, L/\lambda$$

and

$$D = 4\, L/\lambda \qquad (11.14)$$

Thus, in the end-fire mode and with $\xi = 1$, the standard directivity of a linear antenna is twice as great as the directivity in the tilted (scanned) beam case, but the scatter factor does not exceed 10% as before.

When ξ exceeds unity, the directivity first increases as compared with its value at $4L/\lambda$ because the main lobe becomes narrower.

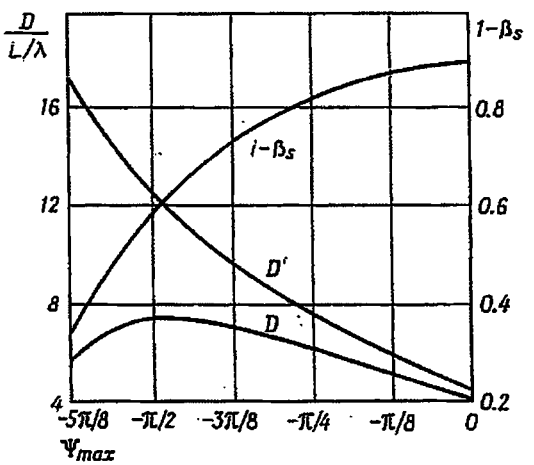


Fig. 11.5. To determination of optimal conditions in end-fire operation

However, the increase in ξ is accompanied by a rise in the sidelobe level, and this leads to a kind of 'saturation' and the subsequent fall in the directivity. The applicable plots of D (Ψ_{max}) and D' (Ψ_{max}) appear in Fig. 11.5 which also shows the $1 - \beta_s$ curve representing the fraction of the radiated power associated with the main lobe of the radiation pattern. As Ψ_{max} varies from zero to $-5\pi/8$, the directivity in terms of the main lobe increases (owing to the fact that the main lobe becomes narrower), but this process is accompanied by an increase in the power scatter factor β_s and a fall in the main-lobe efficiency $(1 - \beta_s)$. An optimum is obtained when

$$\Psi_{max\ opt} \approx -\pi/2$$

where the complete directivity of a long linear antenna is

$$D_{opt} \approx 7.2\, L/\lambda \qquad (11.15)$$

where $\beta_s = 0.43$. In the vicinity of the optimum, the D (Ψ_{max}) plot is blunt, and in practice Ψ_{max} may be chosen to lie somewhat to the right of point $-\pi/2$, say equal to $-3\pi/8$. Then the directivity will fall by a mere 3%, but the scatter factor will be around 0.27 (that is, by 40% lower).

The manner in which the directivity of an ideal linear radiator depends on the relative phase velocity is shown in Fig. 11.6 for two

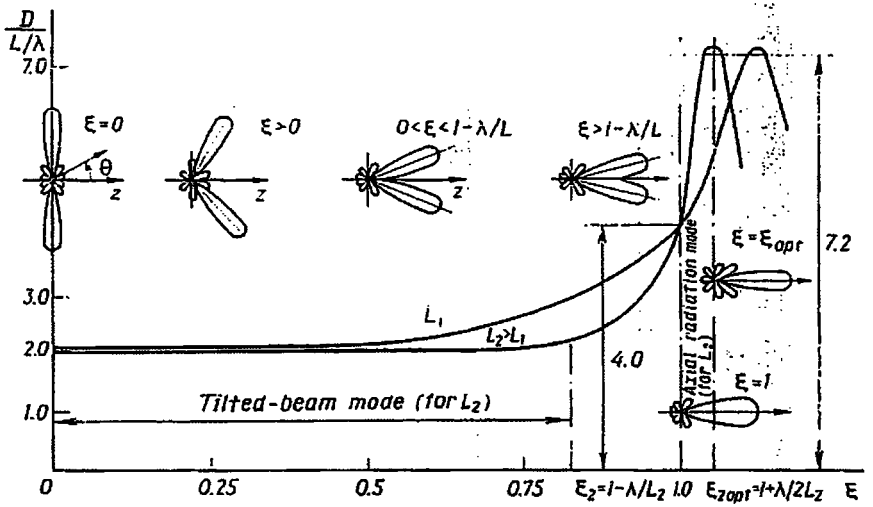


Fig. 11.6. Changes in the directivity and the array factor of a linear antenna with various values of the relative phase velocity ξ

values of L. It is seen that an increase in the antenna length broadens the tilted-beam interval (because of the decrease in the width of the main lobe), and an abrupt reduction takes place in the width of the optimal relative phase velocity interval leading to a maximum directivity in the end-fire condition.

It is to be stressed that all that has been said about the directivity applies to an antenna composed of nondirectional (isotropic) elements. The own directional properties of each element may change both the overall directivity (say, in the case of transverse currents) and its shape (Fig. 11.6). As an example, in a linear antenna with longitudinal currents, the null of radiation of each element is oriented along the axis of the antenna, and the end-fire condition is not feasible in principle.

11.4. The Effect of the Shape of the Amplitude Distribution on the Parameters of a Linear Antenna

Linear antennas with a uniform (or constant) amplitude distribution of the excitation show a rather high sidelobe level (— 13.2 dB), which cannot be tolerated in practice. It has been found, however, that this level can be brought down through the use of a tapered amplitude distribution — a distribution in which the amplitude of the excitation falls off gradually towards the ends of the antenna.

In order to substantiate the above statement, consider a linear antenna of length L in which the excitation (current distribution) is constant in phase and tapering in amplitude as given by the equation

$$I(z) = 1 + \Delta \cos(2\pi z/L), \quad z \leqslant L/2$$

where Δ is the amplitude taper at the ends of the antenna.

On resolving the cosine into the sum of two exponential functions, we reduce the amplitude distribution to the sum of three uniform distributions with a linear phase progression, which are the leading terms of a Fourier series

$$I(z) = \sum_{n=-1}^{1} a_n \exp(jn\xi\beta z), \quad |z| \leqslant L/2 \tag{11.16}$$

where the coefficients are $a_0 = 1$, $a_1 = a_{-1} = \Delta/2$, and the relative phase velocity is $\xi = \lambda/L$.

On the basis of Eq. (11.5), the array factor corresponding to the distribution defined by Eq. (11.16) is likewise a sum of several terms

$$f(\theta) = \sum_{n=-1}^{1} a_n \frac{\sin(\Psi_0 - n\pi)}{\Psi_0 - n\pi} \tag{11.17}$$

where

$$\Psi_0 = 0.5\,\beta L \cos\theta$$

As is seen, the array factor is represented by the leading segment of the Kotelnikov series encountered in many divisions of telecommunications. The array factor in (11.17) has been obtained by what is known in antenna theory as the *partial radiation pattern method.* By this method, each term of the series representing the amplitude-phase distribution of an antenna corresponds to a radiation pattern of its own, or a partial pattern. The partial pattern method provides a better physical insight into how the array factor of a radiating system is shaped and is often used in antenna synthesis, that is, in problems which have as their objective to determine the excitation function of an antenna from the specified radiation pattern.

The summation of the three partial patterns appearing in (11.17) is illustrated in Fig. 11.7 for the case of $\Delta = 0.4$. It is seen that

the addition of two more patterns shifted along the Ψ-axis by $\pm\pi$ and with an amplitude taper of $\Delta/2$ to the main radiation pattern of the $\sin \Psi_0/\Psi_0$ form causes an abrupt decrease in the sidelobe level,

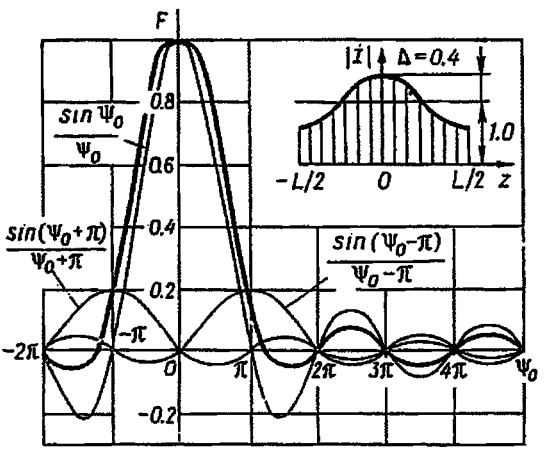


Fig. 11.7. Reduction in the sidelobe level with amplitude taper

accompanied by some broadening of the main lobe. By varying the value of Δ, it is possible to bring down the level of the largest side lobe to a value

$$t\,(\text{dB}) \approx -\,(13 + 13\Delta + 22\Delta^2)$$

Then the half-power beamwidth will be

$$\Delta\theta \approx (1 + 0.636\,\Delta^2)\,51^\circ\,\lambda/L$$

where the term in the brackets is what we call the *beam broadening factor*. For example, for $\Delta = 0.4$ (Fig. 11.7), $t = -22$ dB, and the beamwidth is about $56^\circ\,\lambda/L$, with a beam broadening factor around 1.1.

The broadening of the main lobe causes a reduction in the directivity, the lower bound on which becomes

$$D = D_0 \Big/ \sum_{n=-1}^{n} |\,a_n\,|^2 = D_0/(1 + 0.5\Delta^2) \qquad (11.18)$$

where

$$D_0 = 2L/\lambda$$

is the directivity of a long ideal radiator with a uniform current distribution.

The above analysis of the array factor may be extended to more complex current distributions described by a Fourier series segment containing more than three harmonic terms for which the relative

phase velocity is $n\lambda/L$. In the circumstances, Eq. (11.18), will retain its structure, provided the number of harmonic terms is limited such that

$$|n_{max}| < L/\lambda$$

(the harmonics of the higher orders are almost non-radiating because their main maxima lie in the range of imaginary angles).

Thus, resort to amplitude tapering entails a reduction in the directivity of the antenna, and the penalty for the reduction in the sidelobe level is not only the beam broadening but also a reduction in the directivity. The fractional reduction in directivity D/D_0 is known in antenna theory as the *aperture efficiency*. What is actually meant is the reduction in the effective area of the antenna, which is connected to the directivity by a relation of the form

$$D = 4\pi S_{eff}/\lambda^2$$

The directivity equation (11.18) holds for negative tapers as well, that is, for a form of tapering in which the current distribution increases in amplitude towards the ends of the antenna. In such cases, the phase-shifted correction functions $\sin(\Psi_0 \pm \pi)/(\Psi_0 \pm \pi)$, shown in Fig. 11.7, are subtracted rather than added. It is easy to see that this leads to a substantial increase in the sidelobe level accompanied by a negligible narrowing of the main lobe. Now the directivity decreases because more power goes to the side lobes, that is, because the scatter factor β_s increases. This explains why, in designing an antenna, every effort is taken to avoid inverse amplitude tapering, that is, an increase in the current density towards the ends of the an-

Table 11.1

Amplitude distribution	Array factor	Δ	BBR	SLL	AE	β_s
$(1-\Delta)$	$(1-\Delta)\frac{\sin\Psi}{\Psi}$	0	1.0	−13.2	1.0	0.1
$+\Delta\cos(\pi z/L)$	$+\Delta\frac{2\cos\Psi}{\pi(1-4\Psi^2/\pi^2)}$	0.33	1.12	−20.5	0.93	0.03 max
('cosine at pedestal')		1.0	1.35	−23.5	0.81	0.005 max
$1-(1-\Delta)\,4z^2/L^2$	$\frac{\sin\Psi}{\Psi}+(1-\Delta)$	0.8	1.04	−15.8	0.99	0.07
('parabola at pedestal')	$\times\frac{d^2}{d\Psi^2}\left(\frac{\sin\Psi}{\Psi}\right)$	0.5	1.09	−17.1	0.97	0.03
		0	1.29	−20.6	0.83	0.005 max

Note: Δ = taper; BBR = beam broadening ratio; SLL = sidelobe level; AE = aperture efficiency; β_s = scatter factor

tenna. It also follows from the foregoing that in-phase-driven linear antennas with a uniform (constant) current distribution have the largest directivity as compared with any other smooth distributions. The forms of amplitude tapering and the typical values of the parameters involved are listed in Table 11.1.

Antenna theory also considers various forms of so-called optimal and quasi-optimal amplitude distributions, among them those which provide for the least beam width at a given sidelobe level or for the desired reduction in the sidelobe level on moving away from the main lobe. These distributions are obtained by the methods of antenna synthesis theory and can be found in books on antenna design.

In conclusion it is relevant to note that the parameters of amplitude-tapered in-phase driven antennas remain unchanged in the beam-scanning mode as well, provided the main lobe of the radiation pattern is positioned more than twice or three times of its half-power width away from the axis of the linear antenna.

11.5. The Effect of Phase Errors on the Parameters of a Linear Antenna

Because the distributor is far from perfection in design and also because of unavoidable inaccuracies in the fabrication of antenna elements, the excitation distribution in phase is anything but linear. This has as its results an impairment in the shape of the radiation pattern and a reduction in the directivity, both occurrences falling in the category of phase errors.

Phase errors may be *systematic* (or *deterministic*) and *random*. Systematic phase errors usually have a smooth distribution which can be written as a power series

$$\dot{I}(z) = I(z) \exp j[\Phi(z) - j\xi\beta z]$$

$$\Phi(z) = \Phi_1(2z/L) + \Phi_2(2z/L)^2 + \Phi_3(2z/L)^3 + \ldots$$

where $I(z)$ = amplitude distribution
$\exp(-j\xi\beta z)$ = phase distribution necessary for the scanning mode
$\Phi(z)$ = phase error distribution
Φ_n = constants representing the maximum values of the phase error, raised to the nth power, at the end of the antenna with $z = L/2$

Linear phase errors. The overall linear phase distribution is preserved, but the relative phase velocity is changed

$$\xi_{eq} = \xi - 2\Phi_1/(\beta L)$$

Therefore, in the case of a uniform amplitude distribution the array factor is

$$f(\theta) = \sin(\Psi + \delta\Psi)/(\Psi + \delta\Psi)$$

where

$$\Psi = 0.5\,\beta L\,(\cos\theta - \xi)$$

and

$$\delta\Psi = -\Phi_1$$

Thus, the linear phase error Φ_1 as measured at the end of the antenna relative to its mid-point causes the radiation pattern to shift along the Ψ-axis by an amount equal to $\delta\Psi = -\Phi_1$ (Fig. 11.8). Since the position of the visible range on the Ψ-axis remains unchanged, the

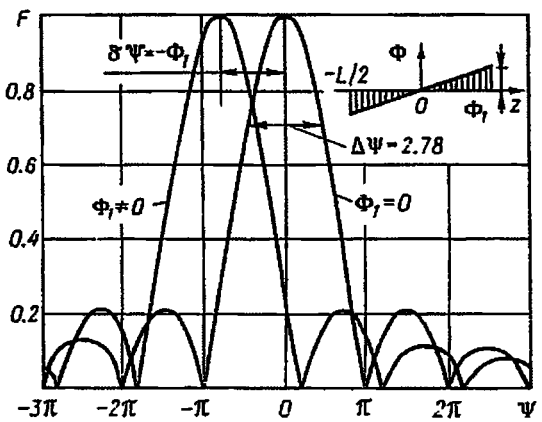


Fig. 11.8. Behaviour of the array factor with linear phase errors in a constant-amplitude antenna

beam deviates in space through some angle $\delta\theta$ towards the antenna's end where the additional phase lag occurs. It is convenient to measure the small parasitic deviation of the maximum relative to the half-power beamwidth, that is, relative to $\Delta\Psi = 2.78$ rad. For long antennas ($L \gg \lambda$), a similar fractional deviation of the beam will take place in terms of observation angles

$$\delta\theta/\Delta\theta = -\Phi_1/2.78 = -\Phi_1^\circ/158^\circ \qquad (11.19)$$

With amplitude tapering, the absolute deviation of the radiation maximum for a fixed Φ_1 remains unchanged, although the relative angular error decreases owing to the beam broadening:

$$\delta\theta/\Delta\theta = -\Phi_1^\circ/(158^\circ K_{bb})$$

where K_{bb} is the beam broadening ratio. The reduction in the relative angular error in the beam position in the case of amplitude tape-

ring is due to the fact that the end portions of the antenna where the phase error is at its greatest play a smaller role.

Quadratic phase errors. In the general case, it is a formidable task to evaluate the integral for the array factor of a linear antenna, defined in Eq. (11.2), in the case of quadratic and cubic phase errors. Preferably, this should be done by numerical methods on a computer,

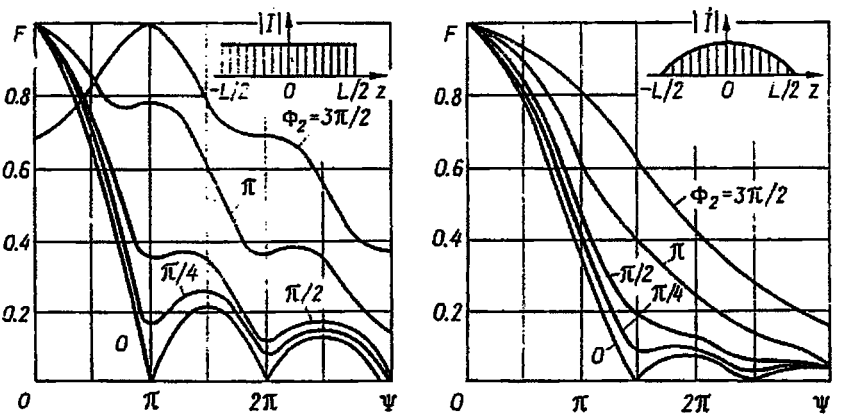


Fig. 11.9. Distortions in the array factor of a linear antenna in the presence of quadratic phase errors

using a fast Fourier transform (FFT) program. The results of such computations are shown in Fig. 11.9. Quadratic phase errors do not cause the maximum of the radiation pattern to shift, but they do cause the main lobe to broaden and the radiation minima (or nulls) to be filled in. Quadratic phase errors are felt most of all in the case of a current distribution of constant amplitude. In such a case, especially when $\Phi_2 > \pi$, the main lobe may even split into two. With amplitude tapering, the antenna ends which are out of phase most of all contribute less to the overall radiated field, and the radiation pattern is thus distorted to a lesser degree.

Quadratic phase errors occur most often in short horn and parabolic reflector antennas, especially when the feed is offset from the focus of the paraboloid in the axial direction. That is why such distortions are often referred to as *focusing errors* or simply *defocusing*. The change in the shape of the radiation pattern caused by quadratic phase errors brings on a reduction in the directivity, that is, a lower utilization of the antenna area. This reduction, in the case of a constant-amplitude current distribution, may be evaluated with the aid of the plot in Fig. 11.10.

Cubic phase errors. In optics, this form of distortion is known as *coma*. The values of the array factor computed for the current distri-

bution of constant amplitude are plotted in Fig. 11.11. Above all, note the shift of the pattern maximum towards the antenna edge which occurs along with an additional phase lag. At small values of Φ_3, the shift of the radiation maximum relative to the beamwidth of an undistorted radiation pattern is given by

$$\delta\theta/\Delta\theta \approx -\Phi_3/1.48\pi = -\Phi_3^\circ/266^\circ \qquad (11.20)$$

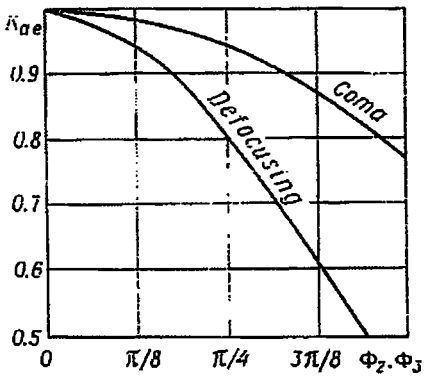


Fig. 11.10. Reduction in the aperture efficiency of a constant-amplitude linear antenna in the presence of quadratic and cubic phase errors

It is seen from comparing Eqs. (11.19) and (11.20) that the deviation of the maximum in the case of a constant-amplitude distribution and a cubic phase error is by a factor 1.7 smaller than it is in the case of a linear error. The distortion in the shape of the radiation pattern consists in that the main lobe of the radiation pattern broadens and becomes asymmetrical, while the side lobes increase in size on one side of the main lobe and decrease on the other, with the increase occurring on the side in which the direction of the main lobe is shifted. The accompanying reduction in the directivity of the antenna can be assessed by reference to the plot of the aperture efficiency as a function of Φ_3 (see Fig. 11.10). With amplitude tapering, the cubic phase error has a lower effect on the shape of the radiation pattern

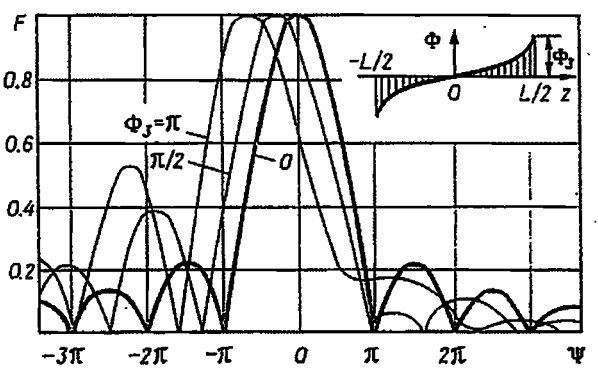


Fig. 11.11. Distortions in the array factor of a linear antenna in the presence of cubic phase errors

because the contribution to the radiated field from the defocused marginal areas of the antenna is reduced.

Cubic phase errors are encountered, for example, in parabolic reflector antennas with the feed offset normal to the focal line (along with a linear change in phase over the aperture).

Random phase errors. Suppose that the current distribution in amplitude and phase in an individual realization of an antenna has the form

$$\dot{I}(z) = I_0 \exp j[\Phi(z) - \xi\beta z] \qquad (11.21)$$

where $\Phi(z)$ is a stationary random process with a mean value of zero and a variance $\overline{\Phi^2}(z) = \alpha$. At low values of variance, the structural function, that is, the mean square of the difference in phase between points z_1 and z_2 may be approximate by the following relation:

$$K(z_1, z_2) = \overline{[\Phi(z_1) - \Phi(z_2)]^2} = \alpha[1 - \exp(-\Delta z^2/\rho^2)] \qquad (11.22)$$

where ρ is the *correlation radius* (or length) and $\Delta z = z_2 - z_1$. The mean square of the difference in phase between points spaced the distance ρ apart is around $0.79\ (\alpha)^{1/2}$ rad. This means that the random phase errors at these points are almost uncorrelated.

For each realization of the distribution defined by (11.21) there is a certain realization of the array factor of the antenna. The dependence of the average parameters of a linear antenna on the phase error variance α and the correlation radius ρ can be established by averaging over a great number of realizations. The following characteristics are most interesting: the shape of the average power radiation pattern, the average directivity, and the most probable sidelobe level.

At low values of phase errors such that $\alpha \ll 1$ and $\rho \ll L$, the average power radiation pattern of a linear antenna with a constant-amplitude distribution is given by

$$\overline{|f(\Psi)|^2} \approx \exp(-\alpha)\left[\left(\frac{\sin\Psi}{\Psi}\right)^2 + \pi^{1/2}\alpha(\rho/L)\exp(-\Psi^2\rho^2/L^2)\right]$$

where

$$\Psi = 0.5\,\beta L(\cos\theta - \xi)$$

With α tending to zero (or with ρ tending to zero), the pattern factor tends to an undistorted shape defined by $\sin\Psi/\Psi$. Phase errors bring about a reduction in the main maximum, or the term $\exp(-\alpha)$, and lead to the superposition of an additional distributed component due to sidelobe radiation on the undistorted radiation pattern. The distributed component causes the radiation nulls to be filled in and the sidelobe level to rise. At $\rho \ll L$, the additional distributed component is almost non-directional, gradually decaying on either side of the main maximum of radiation.

So long as the product $\alpha\rho$ remains constant, *the sidelobe component will be smaller for longer antennas.* For insight into this condition, let us break down the antenna into segments, each of length ρ. Then their number will be $N = L/\rho$. The additional sidelobe component arises as the radiations from the individual segments combine together with random phase shifts. As a result, the power flux density of the sidelobe component is proportional to the number N, and its angular distribution follows the shape of the radiation pattern of a single segment, which means that it is almost nondirectional at $\rho \leqslant \lambda$. In the main lobe, the radiations from the individual segments combine together mostly in phase. This means that the power flux density of the deterministic component of radiation is proportional to N^2, and the additional distributed component is nearly unnoticeable. Thus, as measured in the direction of the main lobe, the sidelobe component is equal to $1/N$ in magnitude or, which is the same, it is inversely proportional to the length of the antenna.

The additional sidelobe component brings down the directivity because of an increase in the scatter ratio. The directivity is reduced most of all when the correlation radius is

$$\rho_0 = 0.3\lambda\,[\ln(2L/\lambda)]^{1/2}$$

and the following estimate holds:

$$D = D_0 \exp(-\alpha) \approx D_0 (1 - \alpha) \tag{11.23}$$

where

$$D_0 = 2L/\lambda$$

is the directivity in the absence of random errors.

In accordance with Eq. (11.23), a reduction of 10% in the directivity will take place when the standard deviation is

$$\overline{\Phi} = \alpha^{1/2} = 0.105^{1/2} = 0.324 \text{ rad} = 18.5^\circ$$

The directivity of an antenna is not too sensitive towards the correlation radius. With $\alpha = 0.105$, the area efficiency rises to 0.916 (instead of the minimal value of 0.900) when $\rho_1 = \lambda/\pi$ and $\rho_2 = L/10$. Therefore, Eq. (11.23) may be used when the correlation radius ranges between 0.3λ and $0.1L$. When the correlation radius tends to zero, the additional sidelobe component disappears, the shape of the radiation pattern is no longer distorted, and aperture efficiency tends to unity. As the correlation radius changes from low values, $\rho < 0.1L$, to high values, the average radiation pattern is distorted in a different manner, namely the main lobe broadens and the sidelobe level is lowered. With the correlation radius greater than or nearly equal to L, the difference in phase between the individual sections of the antenna, as follows from (11.22), decreases, and with the correlation radius tending to infinity the individual re-

alizations of the random radiation patterns retain an undistorted shape, differing from one another by an insignificant phase factor; aperture efficiency likewise tends to unity. Fluctuations in the directivity from one realization of the current distribution to another are negligible at any values of the correlation radius, and so the directivity of any realization in the ensemble is practically the same as the ensemble average. The same reasoning applies to the average beamwidth in which case the following relation holds:

$$K_{bb} = 1 + 1.17\alpha\rho/L$$

where K_{bb} is the beam broadening ratio.

The most unpleasant consequence of random phase errors is the rise in the sidelobe level. The average level of the side lobes closest to the main lobe is given by

$$\overline{f_s^2} = f_{s0}^2 + \pi^{1/2}\alpha\rho/L \tag{11.24}$$

where f_{s0}^2 = power level of a side lobe in the absence of phase errors
$\overline{f_s^2}$ = average level (over many realizations) of the same side lobe in the presence of phase errors

The small excess term $\pi^{1/2}\alpha\rho/L$ added to the original level does not guarantee that in an individual realization of the antenna the level of the lobe in question will not exceed the specified limits. Mathematical analysis shows that with $\alpha \ll 1$ and $\rho \ll L$, the magnitude of the radiated sidelobe field has a modified Rayleigh distribution with a univariate probability density given by

$$w(f_s) = (f_s/\sigma^2)\exp\left(-\frac{f_s^2+f_{s0}^2}{2\sigma^2}\right) I_0\left(\frac{f_s f_{s0}}{\sigma^2}\right) \tag{11.25}$$

where $f_{s0} = |\sin\Psi/\Psi|$ = deterministic component of the sidelobe radiation
$I_0(x)$ = modified Bessel function of order zero
$\sigma^2 = \pi^{1/2}\alpha\rho/2L$ = parameter of the distribution law

Plots of $w(f_s)$ for several values of the ratio f_{s0}/σ are shown in Fig. 11.12*a*. With f_{s0} tending to zero, the distribution function reduces to the usual Rayleigh distribution, whereas at high values of the ratio f_{s0}/σ it tends to the Gaussian distribution. The probability of f_s being smaller than the specified value t is given by the following integral:

$$P\{f_s < t\} = \int_0^t w(f_s)\,df_s$$

The integration results are given in Fig. 11.12*b*. It is seen that with a probability of almost unity the sidelobe level does not exceed the value given by

$$f_{s\max} = f_{s0} + 3\sigma = f_{s0} + 3(\pi^{1/2}\alpha\rho/2L)^{1/2}$$

For probabilities less than unity, the sidelobe level can be evaluated by reference to the plots of Fig. 11.12*b*.

Example 11.1. Let $\alpha = 0.1$, $\rho = \lambda$, and $L = 10\lambda$. Then the mean power sidelobe level in the case of a constant amplitude distribution will, in accordance with Eq. (11.24), be given by

$$\overline{f_s^2} = 0.21^2 + \pi^{1/2} \times 0.1 \times 0.1 = 0.066 \text{ (or } -11.8 \text{ dB)}$$

The parameter of the modified Rayleigh distribution is

$$\sigma = (\pi^{1/2}\alpha\rho/2L)^{1/2} = 0.094$$

and the maximum value of the first side lobe in any realization will not exceed

$$f_{s\,\max} = 0.21 + 3\sigma = 0.50$$

For the probability

$$P\{f_s < t\} = 0.8$$

we find from the plot for $f_{s0}/\sigma = 2$ that

$$t = 0.21 + 0.094 \approx 0.3 \text{ (or } -10.5 \text{ dB)}$$

The phase error variance and the correlation radius should be determined by analysing the distributor type used by the antenna. Assuming the allowable sidelobe level and the probability of its reali-

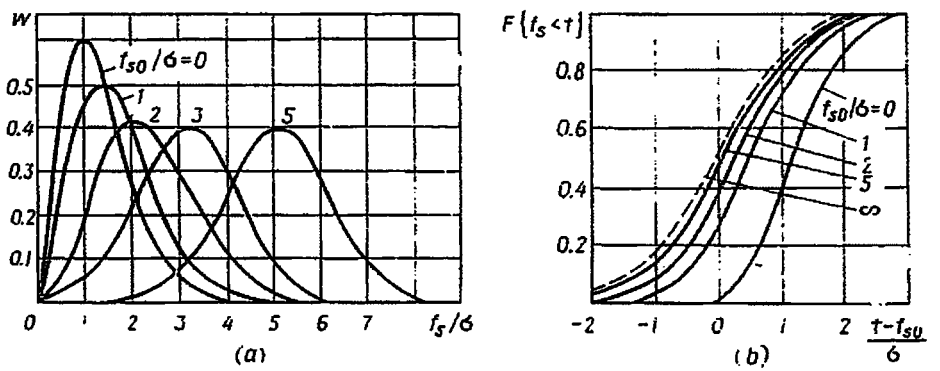


Fig. 11.12. To statistical estimation of the sidelobe level

zation, it is possible to specify the accuracy with which the desired phase distribution of the excitation should be maintained. The equations defining the average radiation pattern, the average directivity and the sidelobe level remain valid even in the presence of small random errors in the amplitude distribution of the excitation although they somewhat increase the overall error variance α which ought not to exceed 0.2 or 0.3.

11.6. Analysis of the Array Factor of an Equally Spaced Linear Array. Suppression of the Grating Lobes

In an equally, or uniformly, spaced linear array, the identical elements (radiators) are spaced the same distance, or spacing, d, apart so that they are positioned at points

$$z_n = (n-1)\,d$$

where n is the nth element (or radiator) (Fig. 11.13), and are driven by currents of a constant amplitude in progressive phase,

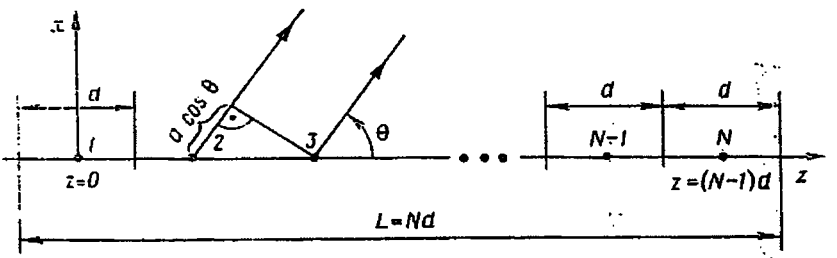


Fig. 11.13. Uniform linear array

that is, the phase changes by a constant angle between adjacent elements along the array

$$\dot{I}_n = I_0 \exp\,[-j(n-1)\Delta\Phi]$$

where $\Delta\Phi$ is the difference in phase between adjacent array elements.

The above radiating system is a discrete analog of an ideal linear radiator. The relative phase velocity of the excitation is given by

$$\xi = \Delta\Phi/\beta d$$

where $\beta = 2\pi/\lambda$.

As before, the case of $\xi = 0$ corresponds to an in-phase excitation, $|\xi| < 1$, to an excitation with a phase velocity exceeding the velocity of light, $|\xi| = 1$, to an excitation with a phase velocity equal to that of light, and $|\xi| > 1$, to an excitation with a slowed-down phase velocity.

According to (11.1), the expression for the array factor of a uniformly spaced array reduces to a geometric progression

$$\dot{f}_N(\theta) = \sum_{n=1}^{N} I_0 \exp\,[j(n-1)\beta d(\cos\theta - \xi)] = I_0 \sum_{n=1}^{N} q^{n-1} = I_0 \frac{1-\dot{q}^N}{1-\dot{q}}$$

where

$$q = \exp [j\beta d (\cos \theta - \xi)]$$

After some identity transformations (including the transfer of the origin of coordinates to the phase centre which coincides with the mid-point of the array), we obtain the following expression for the normalized array factor:

$$F_N(\Psi) = \frac{\sin \Psi}{N \sin (\Psi/N)} \tag{11.26}$$

where

$$\Psi = 0.5N\beta d (\cos \theta - \xi)$$

If we let the number of elements increase to infinity while keeping the array length constant, $Nd = \text{const}$, then $\sin (\Psi/N)$ will tend to Ψ/N, and Eq. (11.26) will automatically be transformed to Eq. (11.5) for an ideal linear radiator. That is why at high values of N and low values of Ψ (such as when Ψ/N is less than $\pi/8$), the functions in (11.26) and (11.5) behave in practically the same manner. This means that the shape of the main lobe and the behaviour of the closest side lobes for a uniformly spaced array are practically the same as they are for a continuous linear radiator of the same length. It then follows that the beamwidth estimates given by (11.7), (11.8) and (11.10) are applicable to a uniformly spaced linear array as well.

The difference in behaviour between the array factors of a continuous linear antenna and of an array arises from the fact that the array factor of an array is a periodic function with a period $N\pi$, and $F_N(\Psi)$ has a maximum of unity level when $\Psi_M/N = M\pi$, where $M = \pm 1, \pm 2, \ldots$, that is, when the denominator turns to zero. This means that in addition to the valid main lobe with a maximum at $\Psi_0 = 0$ the radiation pattern of an array displays unwanted major lobes of order M at point $\Psi_M = MN\pi$. The manner in which these unwanted *or grating lobes* are shaped is illustrated in Fig. 11.14. This diagram is constructed on the same principle as that in Fig. 11.3 for an ideal linear radiator. To make it more instructive, the behaviour of the numerator and denominator of the function $|\sin \Psi/\sin (\Psi/N)|$ is shown separately at the top of the figure.

The plot of $\Psi(\theta)$ shown at the bottom of Fig. 11.14 explains conversion from the generalized angle variable Ψ to real angles θ. The plot also shows the visible range bounded by $\Psi_{max} = 0.5N\beta d (1 - \xi)$ and $\Psi_{min} = -0.5 N\beta d (1 + \xi)$. The total extent of the visible range is determined by the *equivalent length of the array*, $L = Nd$, while the position of the visible range along the Ψ axis can be adjusted by varying the phase shift $\Delta\Phi$ which in turn determines the relative phase velocity. As for an ideal linear antenna, the direction of the main lobe is given by $\cos \theta_0 = \xi$. As is seen from Fig. 11.14,

in addition to the main lobe of the function $F_N(\Psi)$, the visible range encloses two grating lobes. In practice, such a condition cannot be tolerated because the distribution of the radiated power among several major lobes would impair the directivity, thus entailing ambiguity in target bearings and a degradation in noise immunity.

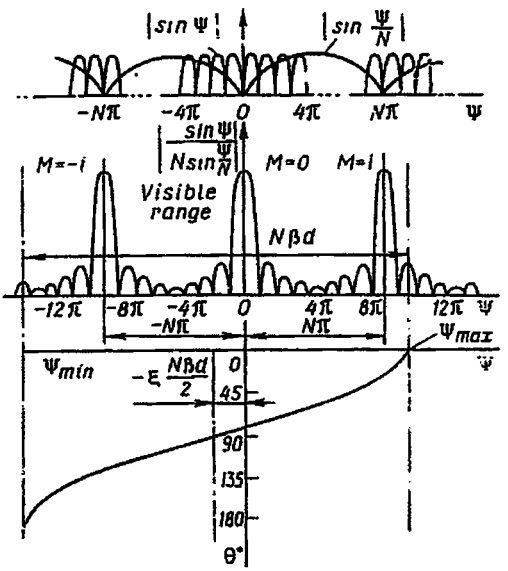


Fig. 11.14. To analysis of the array factor of an array

Let us discuss how the grating lobes can be suppressed by limiting the element spacing in the array, by using directional elements, and by resort to unequal spacing.

Limiting the element spacing. Figure 11.15 shows the generalized angle variable Ψ between the limits $-(N-1)\pi$ and $(N-1)\pi$. Within this interval, no grating lobes are produced, and the side-lobe level does not exceed that of the first side lobe closest to the main lobe. If, with a specified number of array elements, the visible range remains within the limits

$$\Psi_{min} = 0.5N\beta d\,(-1-\xi) > -(N-1)\,\pi$$

and

$$\Psi_{max} = 0.5N\beta d\,(1-\xi) < (N-1)\,\pi$$

no grating lobes can appear within the range of real angles. The two inequalities are equivalent to one condition

$$d \leqslant \left(\frac{N-1}{N}\right)\frac{\lambda}{1+|\xi|} \qquad (11.27)$$

Hence, for a broadside array ($\xi = 0$), the allowable element spacing is

$$d_{max} = (N - 1)\,\lambda/N$$

or somewhat shorter than the wavelength. If the beam scans within a sector from θ_0 to $\pi - \theta_0$, the relative phase velocity varies from $-\cos\theta_0$ to $\cos\theta_0$, and the allowable element spacing is reduced to

$$d_{max} = \left(\frac{N-1}{N}\right)\frac{\lambda}{1+|\cos\theta_0|}$$

In the case of an end-fire array ($|\xi| \geqslant 1$), the allowable inter-element spacing should, in accordance with Eq. (11.27), be less than a half-wavelength.

Sometimes the angle variable Ψ is allowed to vary between $-N\pi/2$ and $N\pi/2$. In such a case, it is guaranteed that the sidelobe

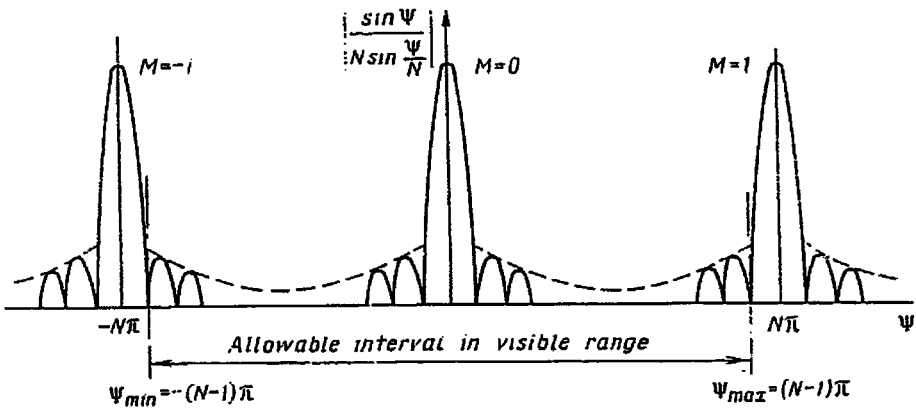


Fig. 11.15. To derivation of the condition for suppression of grating lobes (multilobe occurrence)

level will decrease on moving away on both sides of the main lobe. Then the constraint on the inter-element spacing takes the form

$$d < \lambda/[2\,(1 + |\xi|)]$$

so that the inter-element spacing for a broadside array should be a half-wavelength and for an end-fire array, a quarter-wavelength.

Thus, on satisfying the condition defined in (11.27), the radiation pattern of an array will differ from that of a continuous linear antenna of length $L = Nd$ solely in the behaviour of the far-out side lobes (which have a somewhat higher level in arrays) and the two types of radiating systems may be deemed practically equivalent. The equivalence is preserved with other amplitude and phase distributions, therefore all findings as regards the effect of the amplitude distribution and phase errors in a linear antenna on the shape of the radiation pattern are valid for arrays as well.

Use of directional elements. As follows from the product pattern (or pattern multiplication) theorem, the complete radiation pattern of an array is the product of the *element factor*, or the radiation pattern of one single element, by the pattern of the array, or the *array*

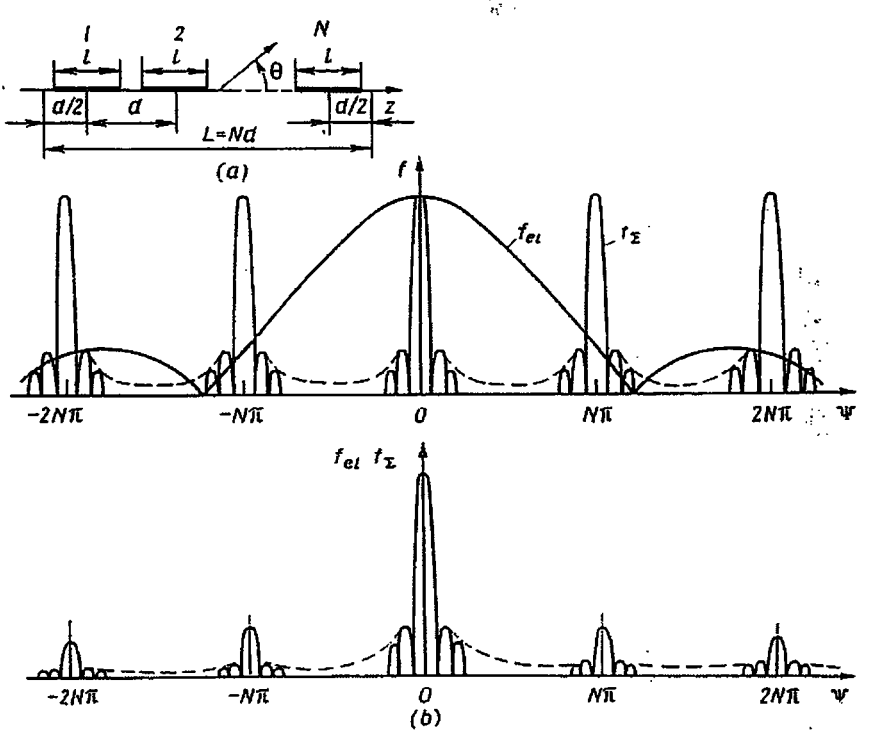


Fig. 11.16. Suppression of grating lobes through the use of directional elements

factor. If one single element radiates only slightly in the direction of the grating lobe of the array, this grating lobe will be suppressed.

Example 11.2. Suppose there is an N-element constant-phase, equal-amplitude array with an arbitrary element spacing d, which consists of identical ideal radiators of length l (Fig. 11.16*a*). Its complete radiation pattern will be

$$f(\theta) = f_{el}(\theta)\, f_\Sigma(\theta) = \left| \frac{\sin(0.5\beta l \cos\theta)}{0.5\beta l \cos\theta} \, \frac{\sin(0.5N\beta d \cos\theta)}{N \sin(0.5\beta d \cos\theta)} \right|$$

Introduce the generalized angle variable

$$\Psi(\theta) = 0.5N\beta d \cos\theta$$

Then,

$$0.5\beta l \cos\theta = l\Psi(\theta)/Nd$$

and the radiation pattern of the array takes the form

$$f(\Psi) = \left| \frac{\sin (l\Psi/Nd)}{l\Psi/Nd} \right| \times \left| \frac{\sin \Psi}{N \sin (\Psi/N)} \right|$$

The respective cofactors are plotted in Fig. 11.16*b*. Let us require that the first grating lobe at $\Psi_1 = N\pi$ in the second factor be suppressed by the first factor to a level of 0.21, which is the standard side-lobe level for a constant-amplitude linear antenna. Since

$$\sin x/x = 0.21$$

at $x = 2.57$, we obtain the condition

$$l\Psi_1/Nd = l\pi/d = 2.57$$

or

$$l = 0.82d$$

The radiation pattern of the array for this value of l is shown in Fig. 11.16*c*. If, however, we choose $l = d$, the nulls of the first factor will be located exactly at the positions of the grating lobes of the second factor, and these grating lobes will be fully suppressed all along the Ψ axis. But this result is trivial because it corresponds to a continuous antenna.

The use of directional elements in an array limits the angle of scan for a linear array. In the case shown in Fig. 11.16, the attempt to effect scanning by introducing phase shifts between adjacent radiators causes the $f_\Sigma(\Psi)$ curve to be shifted relative to the $f_{el}(\Psi)$ curve — a fact which causes a reduction in the principal maximum and a rise in the grating lobe on the side opposite to the direction of beam deflection.

Unequal spacing of radiators. The occurrence of grating lobes in equally spaced arrays of low density is explained by the fact that the waves radiated by any pair of adjacent elements can be combined in phase in several directions, θ_M, where the sum of the spatial path length difference $\beta d \cos \theta$ and the phase difference between adjacent radiators is equal to zero or is a multiple of 2π, that is,

$$\beta d \cos \theta_M - \Delta\Phi = 2\pi M$$

where $M = 0, \pm 1, \pm 2, \ldots$. Hence the likely directions of radiation maximum can be found by the equation

$$\cos \theta_M = (2\pi M + \Delta\Phi)/\beta d = M\lambda/d + \xi$$

A remarkable property of the main lobe in an array with $M = 0$ is the fact that its pointing direction is independent of the element spacing and is solely decided by the relative phase velocity. In contrast, the pointing directions of grating lobes depend on the spacing d/λ in a substantial manner. If the inter-element spacing be made unequal while the relative phase velocity of the exci-

tation, $\xi = \Delta\Phi/\beta d$, unaltered (which can be done by varying $\Delta\Phi$ in synchronism with variations in d/λ) the pointing direction of the major lobe for any pair of elements will remain unchanged whereas the pointing direction of grating lobes will be different for different pairs of adjacent elements so that the grating lobes will be spread out over a wide range of angles.

This idea is embodied in the *unequal* (or *irregular*) *spacing* of elements in an array — a technique alternatively called *density tapering* or *space tapering*. Here, the inter-element spacing varies in a manner other than periodic. The radiation pattern of a density tapered array should be computed by the original equation (11.1) because no general summation technique can be used in a case where the element positions are chosen in an arbitrary manner. The determination of optimal positions for array elements such as would cause grating lobes to spread out is a formidable task which can only be tackled on a computer, using numerical methods. Note that arrays of a lower element density (or "thinned" arrays) than is dictated by the inequality (11.27) have a substantially lower directivity because a good deal of power is wasted in grating lobes. This drawback exists in unequally spaced (or nonuniform) arrays of low density as well.

11.7. The Directivity of a Linear Array

In view of the equivalence between a linear array and a continuous linear antenna, it is quite legitimate to evaluate the directivity of linear arrays by the equations applicable to continuous linear antennas on replacing the length L by an equivalent quantity, Nd. In this way, the following equations for the directivity of arrays can be derived from Eqs. (11.13) through (11.15):

$$D \approx 2Nd/\lambda \text{ and } |\xi| \leqslant 1 - \lambda/Nd$$

in the case of tilted-beam and broadside conditions,

$$D = 4Nd/\lambda$$

in the case of end-fire operation with $|\xi| = 1$, and

$$D_{opt} \approx 7.2Nd/\lambda \text{ and } |\xi_{opt}| = 1 + \lambda/2Nd$$

in the case of an end-fire array with an optimally slowed-down phase velocity of the excitation current.

Obviously, for the above equations to be applicable, it is essential to satisfy the inequality (11.27) which guarantees freedom from unwanted maxima in the visible range.

Along with the above estimates of the directivity of linear arrays, they fully obey all the estimates for the aperture efficiency in the

case of an unequal amplitude distribution and in the presence of phase errors.

It is interesting to elucidate how the directivity of an array depends on the element spacing d/λ. As an example, let us examine the case of an in-phase array with a great number of elements (with N tending to infinity). With d/λ less than unity, only one major lobe exists in the visible range. With d/λ ranging between unity and two,

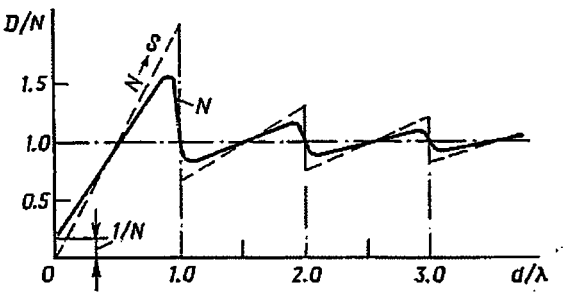


Fig. 11.17. Directivity of an array excited equally and in phase

this lobe is accompanied by two more symmetrical first-order grating lobes. With d/λ ranging between two and three, there appear second-order grating lobes. As one more pair of unwanted maxima occurs in the visible range, the radiated power is shared among the progressively greater number of lobes of unity level, and the directivity of the principal maximum takes the form

$$D = \frac{2N\,(d/\lambda)}{2E\,(d/\lambda)+1}$$

where $E\,(x)$ is the integral part of x. The manner in which the function D/N depends on the element spacing in an array is shown by the dashed curve in Fig.11.17. The full line in the same figure illustrates the behaviour of the same function with a great, but finite number of elements. In the latter case, the grating lobes have a finite width, and they enter the visible range with an increase in d/λ gradually rather than abruptly (as is the case with N tending to infinity).

In the limit of great values of d/λ, the directivity of the array becomes equal to the number of isotropic elements. This can be reasoned out in a non-rigorous manner as follows. Let one single element energized by an excitation current of unity and radiating a power of unity magnitude set up a far-zone field intensity of unity. If the same power be equally divided among the in-phase elements, the current I_0 in each of them will be $1/N^{1/2}$, and the total field intensity in the pointing direction of the main lobe will be multiplied by N to become equal to $N^{1/2}$, which gives $D = N$. From the shape

of the plot in Fig. 11.17 it is seen that this assumption is satisfied approximately at high values of d/λ (in excess of 1.5 to 2.0) and exactly at some points on the d/λ axis, when the element spacing is a multiple of $\lambda/2$. At smaller values of the element spacing, the statement as regards the directivity does not hold any longer, and the increase in directivity reduces to nil as d/λ tends to zero. The cause for this reduction in the directivity lies in *mutual coupling between the elements* which increases as the element spacing decreases and which automatically brings down the current in each element of an in-phase array, although the radiated and input powers remain unchanged. The mutual coupling between the elements of an array does not permit their treatment as independent loads. The input ports of an array are adequately described by the normalized impedance matrix Z in which the diagonal elements are the self-impedances of the radiators and the off-diagonal elements $\dot{z}_{nm}$ are the mutual impedances. When designing an array, the self and mutual impedances are found by electrodynamic calculations or by experiment. For simple radiators (such as half-wave dipoles) a statisfactory approximation is provided by the emf method (see Sec. 9.5).

Using the formal definition of the Z matrix [see Eq. (3.8a)] and writing out the product of the nth row by a column of the amplitude-phase current distribution for each port of the array radiators, we get

$$\dot{v}_n = \dot{z}_{n1}\dot{i}_1 + \dot{z}_{n2}\dot{i}_2 + \ldots + \dot{z}_{nn}\dot{i}_n + \ldots + \dot{z}_{nN}\dot{i}_N$$

On dividing the normalized voltage $\dot{v}_n$ by the input current $\dot{i}_n$, we obtain the normalized input impedance of the nth element in the array

$$\dot{z}_n = \dot{v}_n/\dot{i}_n = \sum_{m=1}^{N} \dot{z}_{nm}(\dot{i}_m/\dot{i}_n) \qquad (11.28)$$

As follows from Eq. (11.28), the input impedance of a radiator is the sum of its self-impedance $\dot{z}_{nn}$ and a set of *insertion impedances* due to the other radiators in the array. With an in-phase and equal-amplitude excitation, the input impedance of a radiator is the sum of the self-impedance and $N - 1$ mutual impedances.

As the array elements are spaced closer together, the mutual impedances approach the self-impedances, and there is an increase in the input impedance of each radiator. This increase is responsible for the reduction in input currents (while the input power remains unchanged) and for the reduction in the array directivity to that of one element (the initial portion of the directivity plot in Fig. 11.17). With any other values of d/λ and with a current distribution of va-

riable amplitude and variable phase, the input impedance as defined by Eq. (11.28) may behave in a very intricate manner, thus causing an appreciable input mismatch with an SWR of 3 to 5 or even greater. With N tending to infinity and with no special measures taken to minimize the mutual impedances, the mismatch at some angles of scan may even cause a complete power reflection from the radiator inputs (giving rise to what are known as *nulls* or *gaps* in the radiation produced by an array in scanning). Thus, the mismatch between the elements of an array may limit the sector scanned by the array, so the element coupling must be considered when one designs matching networks.

Mutual coupling manifests itself differently for the close-in and far-out elements of an array, which means that the input impedances of these elements are different and vary in scanning to a different extent. This leads to a distortion in the amplitude and phase distribution of the excitation current among the elements and to other unwanted consequences such as beam position errors, a reduction in aperture efficiency, and a rise in the sidelobe level.

11.8. Input Power and Gain of an Array

The total input power of an array can be found by adding together the powers applied to the inputs of the individual radiators and can, in accordance with Eq. (3.17), be written as

$$P_{tn} = \langle \mathbf{i}^* \mathbf{R} \mathbf{i} \rangle = \sum_{n=1}^{N} \sum_{m=1}^{N} \dot{i}_n^* \dot{i}_m r_{nm} \tag{11.29}$$

where $\mathbf{R}$ is the real part of the impedance matrix $\mathbf{Z}$ which is symmetrical by virtue of reciprocity. The power is affected by the amplitude and phase distribution of the input current $\mathbf{i}$) and by the resistive parts of the self and mutual impedances (the reactive parts of the impedances do not affect the input power).

Let us write the radiation intensity of an array in terms of element parameters. Suppose that the radiation intensity of the nth element (see Fig. 11.1*a*) of a linear array is, according to Eq. (7.25a), given by

$$\dot{\mathscr{E}}_n(R, \theta, \varphi) = \dot{i}_n \left(\frac{G_n r_{nn}}{4\pi}\right)^{1/2} \dot{\mathbf{F}}_n(\theta, \varphi) \exp(j\beta z_n \cos\theta) \exp(-j\beta R)/R$$

$$= \dot{i}_n \dot{\mathbf{f}}_n(\theta, \varphi) \exp(-j\beta R)/R$$

where G_n = gain of the nth element

$\dot{\mathbf{F}}_n(\theta, \varphi)$ = normalized radiation pattern of the nth element in its own (local) coordinate system

z_n = position of the mid-point of the nth element on the z-axis

$\exp(j\beta z_n \cos\theta)$ = factor taking care of the path length difference between the mid-point of the nth element and the origin of the common coordinate system of the array

The sum of radiation intensities from all the array elements in matrix notation has the form

$$\dot{\vec{\mathscr{E}}}_N(R, \theta, \varphi) = \langle \mathbf{if}(\theta, \varphi)\rangle \exp(-j\beta R)/R \qquad (11.30)$$

where $\langle \mathbf{i}$ = row matrix of the excitation current distribution in amplitude and phase

$\mathbf{f}(\theta, \varphi)\rangle$ = column matrix of the element radiation patterns in the common coordinate system

By difinition, the gain of an array is the ratio between the magnitude of the Poynting vector in the far-zone field and the magnitude of the Poynting vector of an ideal isotropic antenna, given the same input power:

$$G_N = |\dot{\vec{\mathscr{E}}}_N|^2 4\pi R^2/P_{in}$$

On substituting here Eqs. (11.30) and (11.29), we obtain the general equation for the gain of an array composed of arbitrary elements

$$G_N(\theta, \varphi) = 4\pi \,|\, \langle \mathbf{if}(\theta, \varphi)\rangle \,|\,^2/\langle \mathbf{i}^*\mathbf{Ri}\rangle \qquad (11.31)$$

If an array is composed of identical elements, the above equation can be simplified by taking out the common factor

$$G_1(\theta, \varphi) = |\, \dot{\mathbf{F}}_{el}(\theta, \varphi) \,|\,^2 G_1$$

Then the expression for the gain of the array will take the form

$$G_N(\theta, \varphi) = r_{11} G_1(\theta, \varphi) \,|\, \langle \mathbf{i} \exp(j\beta z \cos\theta)\rangle \,|\,^2/\langle \mathbf{i}^*\mathbf{Ri}\rangle$$

where $\exp(j\beta z \cos\theta)\rangle$ is the column matrix of phase factors with the path length differences relative to the selected direction of observation. When an array is phased in a specified direction θ_0, the difference in phase between the driving currents in the elements makes up for the phase lag due to the path length differences

$$\Phi_n = -\beta z_n \cos\theta_0$$

and the gain at the maximum of the array radiation pattern is

$$G_N(\theta_0, \varphi) = r_{11} G_1(\theta_0, \varphi) \Big(\sum_{n=1}^{N} |\, \dot{i}_n \,|^2\Big)/\langle \mathbf{i}^*\mathbf{Ri}\rangle \qquad (11.32)$$

With an equal-amplitude excitation and in the absence of mutual coupling between the array elements ($\mathbf{R} = \mathbf{E}$), the numerator and the denominator of (11.32) can readily be evaluated, and the array gain turns out to be equal to $G_1(\theta_0, \varphi)\, N$, or N times the gain

of one single element in the direction in which the array is phased. Mutual coupling between the elements may cause this gain to increase or to decrease.

Thus, in order to calculate the gain of an array made up of arbitrary elements by Eq. (11.32), it is important to know the complex radiation patterns of the elements and the resistive parts of the mutual impedances. These data can be obtained by calculation or by experiment.

11.9. Travelling-Wave Antennas

Travelling-wave antennas operate in the axial radiation mode and are built around slow-wave guiding structures capable of sustaining surface waves. They are fired at one end, and the travelling-wave condition is assured through the proper choice of parameters for the slow-wave structure and also, but very seldom, through the use of matched terminations at the opposite end. The illumination (or excitation) is generated by a wave which, being launched from a single source (the *feed*), propagates across the guiding structure from end to end and thus illuminates the radiating aperture. This explains why such antennas are often called *series-fed antennas*.

A change in frequency can cause a change in the phase velocity through the slow-wave structure and also affect the efficiency and the quality of match with the feed. As a rule, travelling-wave antennas have a relative bandwidth of a few per cent or, sometimes, a few tens of per cent. A feature unique to this class of antennas is the small cross-sectional area of the radiating system. This is why travelling-wave antennas may be flush-mounted or low-silhouette designs, that is, the radiating apperture is flush with the metallic surface of flying craft on which they are mounted, and hence offer little aerodynamic drag.

Dielectric-rod antennas. These antennas, falling in the class of travelling-wave antennas and also known as *polyrod antennas*, are used at frequencies from 2 GHz upwards. Such an antenna is essentially a dielectric rod (sometimes, a dielectric tube), *1*, circular or rectangular in cross-section, which is several wavelengths long and is driven by a circular or rectangular metallic waveguide (Fig. 11.18). If circular in cross section, the dielectric rod is excited to operate in the lowest hybrid slowed-down HE_{11} (see Fig. 1.10). It is convenient to treat the radiating system of an antenna as a continuous distribution of polarization electric currents

$$\dot{\mathbf{J}}^{e}_{pol} = j\omega\,(\varepsilon_a - \varepsilon_0)\,\dot{\mathbf{E}}$$

which are mainly transverse ones and oriented parallel to the feed dipole, *3*, in the guide. In a zero approximation, the amplitude di-

stribution of the polarization currents along the axis of the dielectric rod may be deemed constant, and their phase distribution varying in a linear fashion. The manner in which the relative phase velocity (that is, the ratio of the free-space velocity c to the wave velocity v in the antenna) depends on rod size and parameters is illustrated in the plots of Fig. 11.19. Using these plots, it is possible to select the rod diameter for which the relative phase velocity would be op-

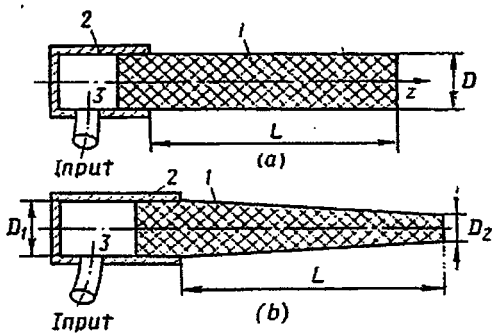


Fig. 11.18. Dielectric-rod antennas: *a*, cylindrical; *b*, tapered

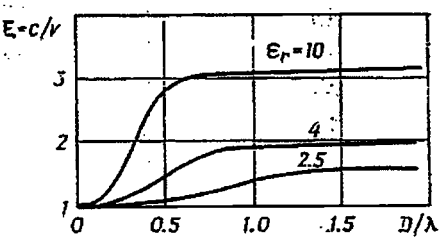


Fig. 11.19. Relative phase velocity for the HE_{11} wave

timal as defined by Eq. (11.19). If the dielectric rod has a dielectric constant (permittivity) of 2 to 5, its diameter will usually range between a half and a third of the wavelength.

The radiation pattern of a dielectric-rod antenna could approximately be calculated by Eq. (11.5) which, however, applies to isotropic-source antennas and ignores the directional properties of an individual antenna element. Nor does it take into consideration the radiation produced by the polarization current wave reflected from the rod end. Since, however, the phase velocity of the wave in the antenna is close to that of light (or the free-space velocity), the reflection coefficient from the rod end is small, being around 0.15. That is why the radiation due to the reflected wave may be neglected.

Figure 11.20 shows experimental radiation patterns of a rectangular polyrod antenna with cross-sectional dimensions of $\lambda/2$ by $\lambda/3$ for three different values of the relative length L/λ : 3, 6, and 9. The same figure includes the measured values of directivity. The width and directivity of the main beam show a satisfactory agreement with estimates:

$$\Delta\theta = (107 \text{ to } 61)^\circ(\lambda/L)^{1/2}$$

and

$$D \approx (4.0 \text{ to } 7.2)\, L/\lambda$$

but the sidelobe level is too high. The fact that there are no nulls of radiation ('gaps') between the lobes is due to the attenuation of the wave as it is propagated through the dielectric rod.

There is one more factor which adversely affects the radiation pattern of the dielectric-rod antenna. What is meant is the spontaneous radiation of the feed at the joint between the metallic guide and the dielectric rod. As the parasitic radiation of low directivity from the feed interferes with that due to the polarization currents in the

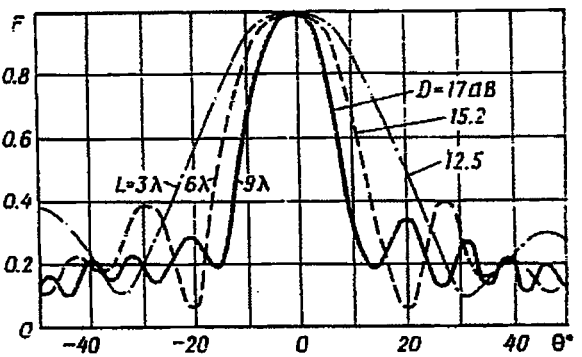


Fig. 11.20. Radiation pattern of a dielectric-rod antenna

rod, the main beam is unduly distorted and the sidelobe radiation gains in strength. If the rod diameter is great in comparison with the wavelength, most of the power is transferred to the slowed-down HE_{11} mode, and only a small fraction is directly radiated at the joint of the two guides. As this happens, however, the relative phase velocity ξ grows too much, the optimal antenna length $L_{opt} = \lambda/2\,(\xi - 1)$ is too small, and there is no way of obtaining a high directivity. One way out is to use a tapered dielectric rod, such as shown in Fig. 11.18*b*. The rod taper causes the relative phase velocity to decrease gradually along the rod. By adjusting the rod taper, it is possible to obtain an optimal value for the average relative phase velocity at any given antenna length.

In a dielectric-rod antenna, the feed dipole in the metallic waveguide should preferably be placed outside the dielectric rod. This improves the efficiency because the higher modes excited by the dipole will die out near the feed point and will not penetrate into the dielectric rod.

Apart from their use as stand-alone units, dielectric-rod antennas are often employed as low-directional elements in phased arrays.

Helical antennas. Also known as 'helical-beam antennas', they may be cylindrical, flat, or conical in shape. At frequencies from 300 MHz upwards, cylindrical helical antennas are used (11.21*a*).

They radiate a circularly polarized field confined mainly to a single lobe located along the axis of the helix.

A cylindrical helical antenna consists of a wire helix *1*, which is several wavelengths long, the turn diameter being equal to λ/π. One end of the helix is left free, and the other end is connected to the inner conductor of a coaxial line, *3*. The outer conductor of the coaxial line is connected to a metal shield, *2*. A travelling electric-cur-

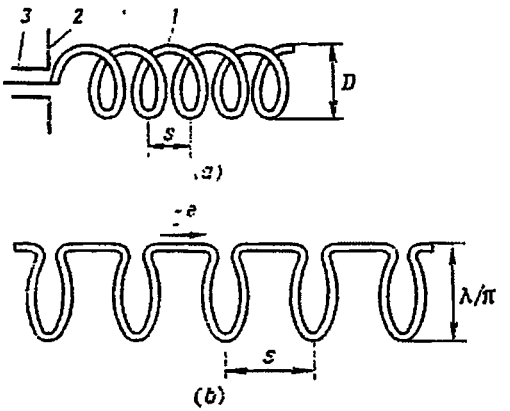


Fig. 11.21. Helical antennas: *a*, general arrangement; *b*, equivalent circuit

rent wave is excited in the helix, and the radiation maximum is oriented along the axis in the direction of wave propagation.

The radiating system may be pictured as consisting of a set of flat circular turns spaced a distance s apart and energized in series (Fig. 11.21*b*). The current distribution in each turn has the form

$$\dot{I}_0 \exp(-j\beta l)$$

where $\dot{I}_0$ is the current at the start of the turn and l is the coordinate reckoned along the turn. With $D = \lambda/\pi$, the turn perimeter will contain exactly one wavelength, therefore the current distribution along each turn may likewise be written as $\dot{I}_0 \exp(-j\varphi)$, where φ is the azimuthal angle in the plane of the turn. The radiation produced by a turn with this type of current distribution has already been discussed in Sec. 10.6, and it is legitimate to use Eqs. (10.3) and Fig. 10.25 in order to analyse the radiation pattern and polarization of one turn.

However, the radiation pattern of a turn may be interpreted in a different way. The current distribution along a turn

$$\dot{I}_0 \exp(-j\beta l) = \dot{I}_0 \cos \beta l - j\dot{I}_0 \sin \beta l$$

is in fact the superposition of two standing waves in phase quadrature with each other, with the amplitude of one varying cosinusoidally and that of the other, sinusoidally (Fig. 11.22). Thus, there are four bent half-wave dipoles. One pair of dipoles which are in phase with each other is oriented along the y-axis and radiates a maximum of power in the direction of the z-axis, with the vector E parallel to the y-axis. The other pair of dipoles which are likewise in phase with each other is oriented in the direction of the x-axis and radiates a maximum of power likewise in the direction of the z-axis, but with

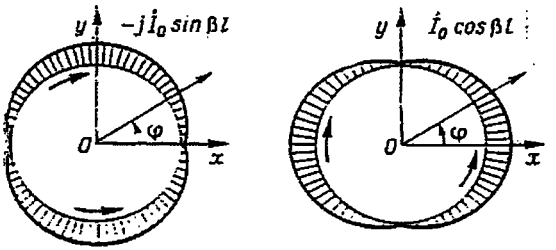


Fig. 11.22. Current distribution in a helical antenna

the vector E parallel to the x-axis. The dipole pairs are energized with a phase difference of 90°, and the field radiated in the z-direction is circularly polarized. At some angle to the z-axis, the field is elliptically polarized, and in the xOy plane, it is linearly polarized.

Because the dipoles in each pair are in phase, their total radiation impedance, considering the mutual coupling between them, turns out to be rather high. The input impedance of each turn is close to the characteristic impedance, and if we recall that adjacent turns operate nearly in phase because s/λ is small, it is an easy matter to realize why a helix of a diameter λ/π is operating in the travelling-wave mode very nearly. The phase velocity of the excitation wave in adjacent turns along the z-axis is somewhat lower than the velocity of light, and the result is the axial-beam mode and an end-fire antenna why a slowed-down phase velocity (Fig. 11.23*b*).

If, on the other hand, the helix diameter D is substantially smaller than λ/π, the currents at diametrically opposite points on a turn flow in opposite directions, and the radiation resistance of the turn is very small (this is what is known as the loop effect). A standing-wave condition now exists in the antenna, the radiation along the axis of the helix is zero, and the maximum of radiation from each turn and from the entire antenna is normal to the helix axis (Fig. 11.23*a*). This condition is known as the normal mode, and the antenna is called a normal-mode helix.

With D substantially greater than λ/π, the currents at opposite points on each turn and in adjacent turns are again out of phase, the travelling-wave condition is upset, and the radiations in the direction of the z-axis from the individual elements of a turn cancel out; the peak of the antenna pattern now makes an angle with the axis (Fig. 11.23*c*).

The principal mode is the *axial radiation* or *axial-beam mode.* Calculations and measurements show that in an end-fire helical antenna the phase velocity is slowed down progressively less with in-

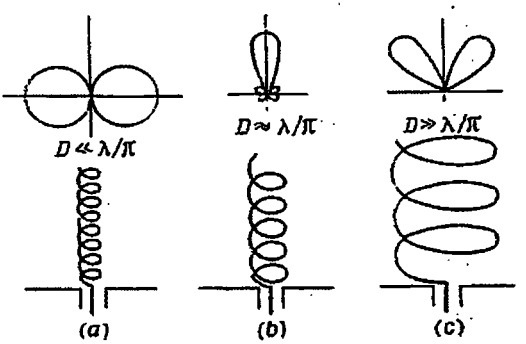


Fig. 11.23. Radiation modes of a helical antenna

creasing frequency. This serves to broaden the bandwidth of the antenna. Thus, for a helical antenna with a pitch angle equal to $\alpha = 12°$ and with a relative turn length equal to $L/\lambda = 0.72$ to 1.2, the relative phase velocity ranges from $\xi = 1.67$ at the lowest frequency to $\xi = 1.1$ at the highest frequency. For the field radiated along the axis of the helix to be circularly polarized, it is essential to satisfy the following condition:

$$\beta\xi L - \beta s = 2\pi$$

or

$$\xi L = s + \lambda \tag{11.33}$$

If, however, we treat the helical antenna in question as an end-fire antenna with a slowed-down phase velocity, then in order to obtain a maximum directivity the phase difference between the radiated fields from the first and last elements of the antenna as seen at an observation point on its axis should be equal to π, which means that the following condition must be satisfied

$$\beta\xi L - \beta s = 2\pi + \pi/n$$

where n is the number of turns in the helix, or

$$\xi L = s + \lambda + \lambda/2n \tag{11.34}$$

Thus, when the condition defined by Eq. (11.33) is satisfied an exactly circular polarization is obtained along the helix axis, whereas when the condition defined by (11.34) is satisfied, the directivity is a maximum when the polarization is slightly elliptical. Because the relative phase velocity ξ increases with increasing frequency, the conditions defined by (11.33) or (11.34) are violated only slightly, and the parameters of the helical antenna remain satisfactory over the wavelength range extending from $0.7\lambda_0$ to $1.2\lambda_0$, where λ_0 is the wavelength for which optimal antenna dimensions have been chosen.

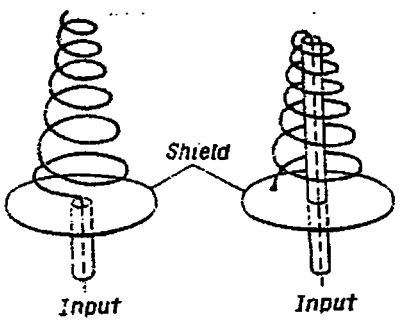


Fig. 11.24. Conical helical antenna

Conical helical antennas are more broadband. Examples of such antennas with their broad base facing the shield are shown in Fig. 11.24. In one case, the antenna is base-fed, and in the other, it is top-fed. Measurements made on conical helical antennas with $n = 10$, $\alpha = 6°$ and $D_{max}/D_{min}=3$ show that the axial radiation of a base-fed antenna remains unchanged over nearly a 2 : 1 bandwidth and that of a top-fed antenna, over nearly a 3 : 1 bandwidth.

Slow-wave impedance antennas. These antennas owe their name to the fact that they are built around impedance structures capable

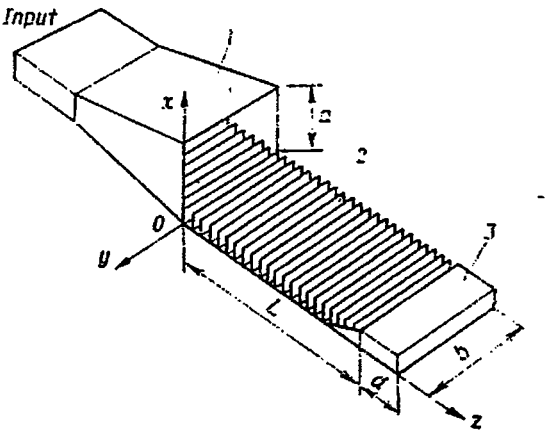


Fig. 11.25. Corrugated-surface antenna

of sustaining surface waves. Figure 11.25 shows an impedance antenna on a flat sheet (ground plane), *3*. The surface wave launched by a small horn (a horn launcher), *1*, is propagated along a corrugated

surface, 2, so that its amplitude remains nearly unaltered because wave reflection from the end of the structure is negligible. The electromagnetic field set up at the surface of the corrugated structure contains the components $\dot{H}_y$ and $\dot{E}_z$ which can, by the equivalence principle, be thought of as corresponding to equivalent surface currents $\dot{J}_z^e$ and $\dot{J}_y^m$ in the yOz plane. These currents are distributed along the z-axis in the same manner as a travelling wave with a phase velocity along the antenna less than the free-space velocity. Hence, the corrugated-surface structure turns out to be equivalent to a slower-than-light travelling-wave radiating system. In order to sustain a slowed-down wave, the surface impedance of the corrugated-surface structure

$$\dot{Z} = \dot{E}_z/\dot{H}_y$$

must be purely inductive. For this condition to be satisfied, the corrugation (slot) depth must be less than a quarter of a wavelength.

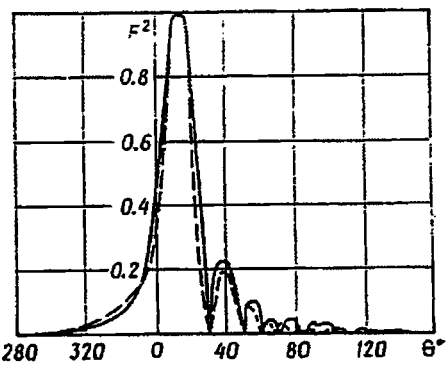


Fig. 11.26. Radiation pattern of a corrugated-surface antenna on a finite screen

When the corrugated-surface structure is matched well to the horn launcher, most of the input power is transferred to the surface wave and the remainder is radiated. Therefore, the radiation of a corrugated-surface antenna would mainly be decided by the magnetic surface currents of the corrugated-surface structure provided that the ground plane is infinite in size, and the radiation pattern would, in a zero approximation, be defined by Eq. (11.5). The directivity along the z-axis would be a maximum at an optimal relative phase velocity given by

$$\xi_{opt} = 1 + \lambda/2L$$

Actually, however, the ground plane has a finite dimension d (Fig. 11.25), and the peak of the radiation pattern is, because of this, 'forced' away from the ground plane. The typical radiation pattern of a corrugated-surface antenna in the xOz plane is shown in Fig. 11.26.

The bandwidth of impedance antennas is mainly determined by the quality of match between the launcher and the radiating surface and is usually 15% to 20%. A horn launcher may be replaced by an assembly of several slot radiators cut in the ground plane and energized in progressive phase corresponding to the velocity of propaga-

tion of surface waves in the structure. This type of antenna is mounted flush with the metallic ground plane.

Multielement Yagi-Uda arrays. Closely coupled parasitic arrays where the parasitic element may operate as either a director or a reflector were first described by S. Uda and then by H. Yagi. They are often referred to as Yagi arrays, although they are also known as Yagi-Uda (or Uda-Yagi) arrays. When used at frequencies between 100 and 1000 MHz, they are ordinarily assembled from rigid metal dipoles (Fig. 11.27). At higher frequencies, it is more convenient to use stripline and microstrip designs.

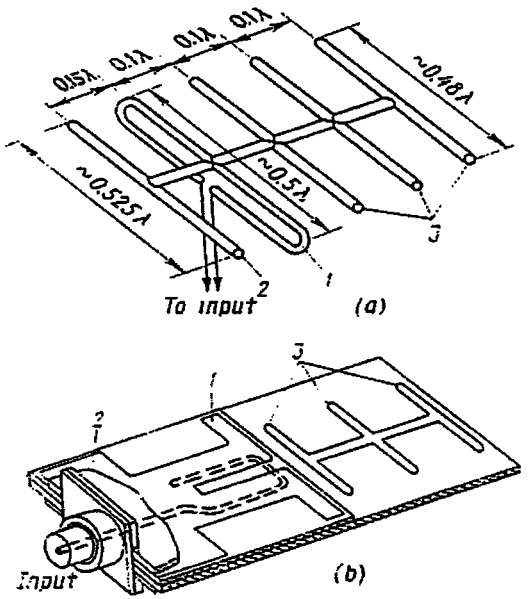


Fig. 11.27. Yagi-Uda antennas: *a*, wire; *b*, stripline

In a Yagi-Uda array, one dipole is active, *1*, and all the other dipoles are passive. One of the passive dipoles, placed back of the active element, acts as a *reflector*, *2*, and the remaining dipoles placed in front of the active dipole operate as *directors*, *3* (see Sec. 9.6). The reflector length is somewhat greater than a half-wavelength, and the director length is somewhat smaller than a half-wavelength. Only one reflector is provided in a Yagi-Uda array because if more were installed they would be only slightly excited and would have a negligible effect on the radiation pattern. Ten or more directors may be used because the antenna radiates towards the directors, and they are therefore excited consecutively rather strongly.

The dipoles of a Yagi-Uda array can be secured at their mid-points to a longitudinal metal rod (Fig. 11.27*a*). This rod is not excited because the electric lines of force, E, cut it at right angles. The active (or driven) dipole is usually of a folded configuration and is secured to the supporting rod at the mid-point of its uncut side. In this way a Yagi-Uda antenna can be built without resort to insulators and grounded for direct current (for the purpose of lightning protection).

For the directivity of a Yagi-Uda array to be a maximum the spacing between and the length of its directors should be carefully selected so as to obtain certain relationships between the currents in the dipoles at a specified frequency. A tuned Yagi-Uda array is

highly sensitive to changes in frequency. That is the reason why Yagi-Uda arrays are narrowband systems — with an allowable twofold change in directivity the bandwidth is a few per cent.

Closely related to Yagi-Uda arrays as regards directional characteristics are *disc-on-rod* antennas (Fig. 11.28). In such antennas the passive metal discs secured to the central rod may be likened to

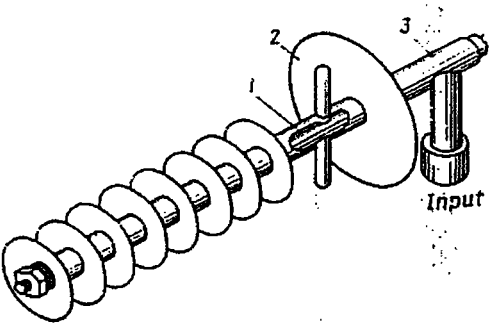


Fig. 11.28. Disc-on-rod antenna: *1*, balun; *2*, shield; *3*, metal insulator

an array of flat dipoles. There is however a different approach in which the discs and the central rod taken together are treated as a cylindrical slow-wave corrugated-surface structure.

Among the advantages offered by Yagi-Uda and disc-on-rod antennas are simple design, high efficiency (close on 100%) because there are no dielectric insulators, and a relatively low weight. The directivity can further be enhanced by arranging several parallel radiators in an array.

11.10. Slotted Waveguide Arrays

This type of antenna is an array of many radiating slots excited by a common waveguide. Such antennas are mostly used on board aircraft to provide a highly directional radiation pattern.

Slot types. Most often, use is made of half-wave resonant slots cut in a broad or a narrow face of a rectangular waveguide operating in the TE_{10} mode.

An isolated slot is excited when it is crossed by the electric currents flowing on the internal surfaces of the guide walls. As the TE_{10} is propagated in a rectangular waveguide, three components of the surface electric current are present there, namely two transverse components $\dot{J}_x$ and $\dot{J}_y$ due to the longitudinal magnetic field component $\dot{H}_z$, and one longitudinal component $\dot{J}_z$ due to the transverse magnetic field component $\dot{H}_x$. The longitudinal current component

exists only on the broad faces of the guide, whereas the transverse current components which are in phase quadrature with it exist on both the broad face ($\dot{J}_x$) and the narrow walls ($\dot{J}_y$) of the guide. Proceeding from the surface current distribution, it is an easy matter to picture the physical excitation pattern for various slot types in a guide (Fig. 11.29).

A type I transverse slot cut in a broad face is excited by the longitudinal currents, its excitation intensity being reduced as the slot

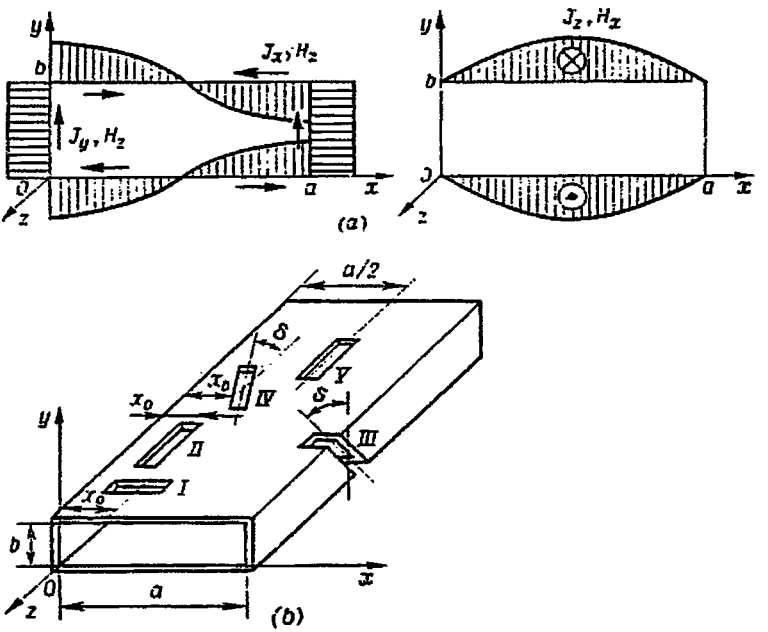


Fig. 11.29. Current distribution in waveguide walls (*a*) and various slot types (*b*)

centre is offset progressively more from the centre of the broad face. A type II longitudinal slot intercepts the transverse current, provided it is offset from the centre of a broad face of the guide. Its radiation goes up as the slot is moved closer to the edge of the guide wall and is totally non-existent when $x_0 = a/2$. When a longitudinal slot is arranged to lie on each side of the centre line of the guide (a centred longitudinal slot), the radiated field undergoes a phase reversal because of the reversal in the direction of the surface current. A type III inclined slot cut in a narrow face of the guide is excited by a transverse current of constant amplitude. Therefore its radiation intensity can be adjusted by varying the angle of tilt δ. At $\delta = 0$

the slot is not excited at all, whereas at $\delta = 90°$ the radiation is at its maximum. An offset inclined slot, type IV, in a broad face intercepts both longitudinal and transverse currents. In consequence, its radiation intensity (and the phase of the radiated field) depends on both the centre coordinate x_0 and the angle of tilt δ.

All of the above slot types radiate a linearly polarized field with the E vector normal to the long dimension of the slot. Where it is desired to have a circularly polarized field, resort is made to cross-shaped slots obtained by cutting type I and II slots so that their centres coincide. The surface currents $\dot{J}_z$ and $\dot{J}_x$ which respectively excite the transverse and the longitudinal slot are in phase quadrature, and this supplies one of the conditions necessary to produce a circularly polarized radiated field. By varying the centre coordinate x_0, it is possible to assure another condition, namely the equality in amplitude between the excitations applied to the mutually perpendicular slots. By moving a cross-shaped slot to the right or left of the guide axis (or by reversing the direction of wave propagation in the guide), it is possible to obtain a right-hand or a left-hand circularly polarized field. Cross-shaped slots may also be made up of two offset inclined slots which have a common centre and are oriented to make an angle of $\delta = 45°$ while being at right angles to each other. Such slots have the best electrical characteristics and offer the best structural possibilities.

Sometimes, slots in a waveguide can be excited by suitable reactive elements. For example, a half-wave longitudinal slot, type V (Fig. 11.29), arranged to lie precisely along the centre line of a broad face, can be excited by a capacitive probe located next to the slot centre and adjustable for immersion depth. The TE_{10} mode field induces in the probe an electric current in the same way as it would do in a receiving dipole. This current extends onto the broad face and spreads radially around the probe. Some of the current crosses the slot and excites it. The direction of the crossing current will be reversed if the probe be placed on the other side of the slot — this will cause the excitation to undergo a phase reversal.

Equivalent circuits of slots. A radiating slot loads the waveguide and affects the conditions existing in it. Some of the power channelled by the guide is radiated by the slot, some is reflected towards the generator (as would be the case with any discontinuity), and the remainder proceeds onwards. When analysing the response of the slot to the waveguide, the latter is replaced by an equivalent long line, and each slot is replaced by an equivalent circuit in the form of a two-port with losses to simulate radiation. This approach permits the use of analytical tools from the theory of microwave circuits for purposes of multislot antenna analysis and design. The equivalent circuits are the simplest of all for resonant longitudinal and resonant transverse slots made in a broad face of a guide.

A *transverse slot* in a broad face intercepts longitudinal current lines, and so its equivalent circuit has the form of a lumped series resistance r in an equivalent line, as shown in Fig. 11.30*a*. The normalized resistance is found by methods borrowed from electrodynamics, proceeding from the power balance in the slotted waveguide and its equivalent circuit. This gives the following approximate equation:

$$r \approx 0.523\ (\lambda_g/\lambda)^3\ (\lambda^2/ab)\ \cos^2\ (\pi\lambda/4a)\ \sin^2\ (\pi x_0/a) \qquad (11.35)$$

where x_0 is the distance from the centre of the slot to the edge of the broad face.

When the guide is shorted by a plunger, a standing-wave condition is produced, with the antinodes of the longitudinal electric currents

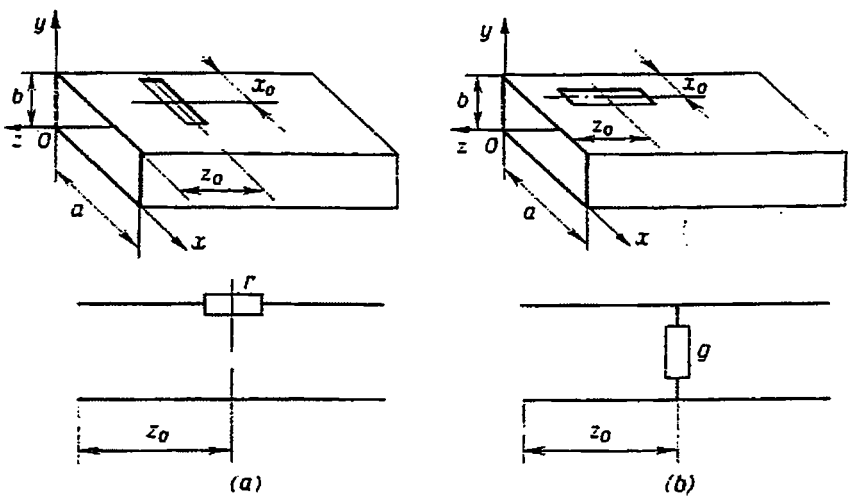


Fig. 11.30. Equivalent circuits of (*a*) a transverse slot and (*b*) a longitudinal slot in a rectangular waveguide

removed from the plunger by distances which are multiples of $\lambda_g/2$. It is at these antinodes that the slots cut in a shorted guide must be positioned for maximum excitation.

A *longitudinal slot* intercepts lines of transverse surface electric currents. In a way, these currents branch off from the conductors of the equivalent two-wire line into parallel-connected shorted quarter-wave stubs. That is why for a longitudinal slot the equivalent circuit has the form of a normalized shunt conductance g (Fig. 11.30*b*) approximately given by the following equation:

$$g \approx 2.09\ (a/b)\ (\lambda_g/\lambda)\ \cos^2\ (\pi\lambda/2\lambda_g)\ \cos^2\ (\pi x_0/a) \qquad (11.36)$$

where x_0 is the distance from the slot to the edge of the broad face.

In a shorted guide, the antinodes of transverse surface currents are removed from the plunger by distances which are given by

$$\lambda_g/4 + m\lambda_g/2, \quad m = 0, 1, 2, \ldots$$

It is at these planes that longitudinal slots must be located for maximum excitation.

Equivalent circuits can be developed for other slot types as well. The element values of these circuits are found by electrodynamic analysis and can be used to design actual arrays.

Resonant slotted waveguide arrays. These units are built around a shorted guide, with the slots spaced exactly λ_g apart if they are

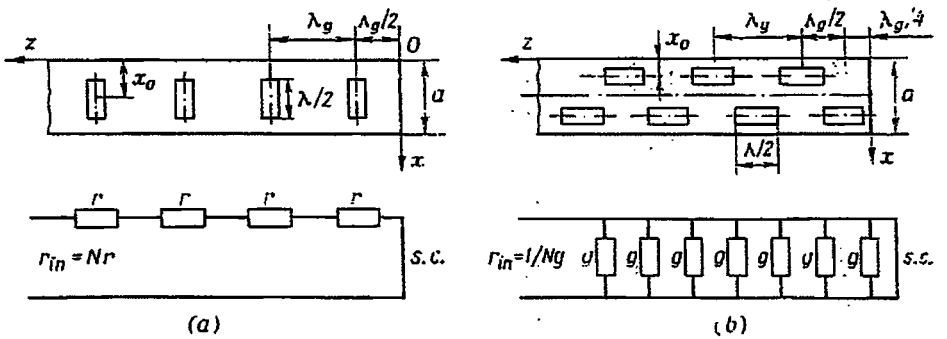


Fig. 11.31. Resonant slotted waveguide arrays: *a*, constant-phase-coupled; *b*, variable-phase-coupled

coupled with the guide field in-phase (Fig. 11.31*a*) or exactly $\lambda_g/2$ apart if they coupled with the guide field with a phase reversal. In each case, all the slots of a resonant array are excited in-phase (even though adjacent longitudinal slots undergo a phase reversal due to their placement on either side of the centre line of the broad face). In consequence, the peak of the radiation pattern is normal to the guide axis, giving broadside radiation. If the values of r or g are the same for all slots, the excitation will be uniformly distributed in amplitude (on neglecting the mutual coupling between the slots over the external space), and the radiation pattern in a longitudinal plane passing through the guide axis will be defined by Eq. (11.26) which is extended to include an additional factor in the form of the radiation pattern of a single slot in a plane screen. The normalized input resistance of a transverse-slot array, as measured at the beginning of the guide, is

$$r_{in} = Nr$$

and that of a longitudinal-slot array,

$$r_{in} = 1/Ng$$

where N is the number of slots. Using the condition for an ideal input match, $Nr = 1$ or $Ng = 1$, it is possible with the aid of equations similar to (11.35) or (11.36), to select slot parameters such that the input SWR will be equal to unity at the resonance frequency of the array. The quality of match and in-phase excitation are preserved over a relative narrow bandwidth not exceeding a few percent.

In transverse-slot antennas the slots are spaced far apart, and this fact is responsible for the occurrence of unwanted grating lobes (multilobe occurrence). They can be avoided, for example, by arranging

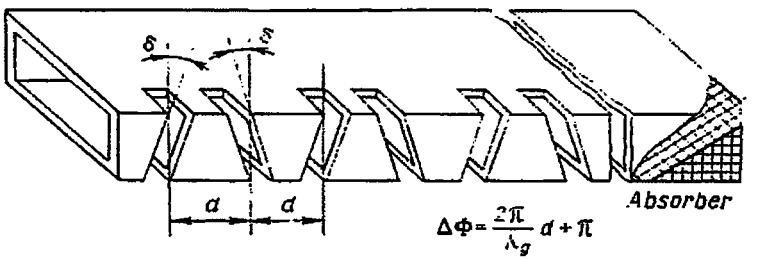


Fig. 11.32. Nonresonant slotted waveguide array

in parallel two in-phase excited identical arrays so that all the slots (and the short-circuiting plunger) in one are shifted through $\lambda_g/2$ relative to the slots in the other. What is thus formed is an equivalent linear array with half as great the slot spacing, and the grating lobes are crowded into the range of imaginary angles.

Nonresonant slotted waveguide arrays. These differ from their resonant counterparts in that the waveguide is terminated in a matched load so that, in the absence of slots, a travelling TE_{10} wave is produced in it. The slot spacing d somewhat differs from $\lambda_g/2$ (for transverse slots, the spacing somewhat differs from λ_g). As a result, the slots are excited in progressive phase, the phase difference being

$$\Delta\Phi = 2\pi d/\lambda_g$$

between adjacent in-phase coupled slots, and

$$\Delta\Phi = 2\pi d/\lambda_g \pm \pi$$

in the case of adjacent phase-reversal coupled slots.

An example of a nonresonant array using inclined slots cut in the narrow dimension of a guide is shown in Fig. 11.32. The slots alternate in their inclination from the vertical, and this arrangement assures that they are excited with phase reversal. If d and $\Delta\Phi$ are given or known in advance, the direction of maximum radiation and the shape of the radiation pattern can be found by Eqs. (11.6) and (11.26). The amount of coupling between the slots and the waveguide is chosen such that a mere 5% to 20% of the input power of the

array reaches the terminal absorber, and the efficiency is 80% to 95%.

Since the slot spacing in nonresonant arrays is other than $\lambda_g/2$, reflections from the individual slots cancel out to a considerable extent, and the input SWR is close to unity over a rather broad bandwidth. It is only at the frequency where $d = \lambda_g/2$ and the reflections add together that the SWR abruptly goes up and the radiation that should be normal to the guide axis is drastically reduced (the so-called 'normal effect'). Therefore, if it is planned to operate an array in the normal mode (that is, with its maximum of radiation normal to the guide axis), each slot should be carefully matched to the waveguide with the aid of a suitable tuning element.

If we wish to preserve the travelling-wave condition along a nonresonant slotted waveguide array and to avoid the undesirable 'image' beam that might arise due to the propagation of the reflected wave, the amount of coupling between the slots and the guide should be kept to a reasonable minimum. This can be achieved through the proper choice of the angle that each slot in the narrow dimension of the guide makes with the vertical or of the positions for the centre points of the slots in the broad dimension. Likewise, by adjusting the amount of coupling between slots and waveguides it is possible to cause the amplitude of the excitation to vary in any desired manner along the antenna.

Slotted waveguide arrays, both resonant and nonresonant, can use various feeding methods and with a great number of slot rows arranged in parallel, make it possible to build highly directional antennas. The key advantages of such antennas may be summed up as follows.

(1) They have no projecting parts and use a compact waveguide distribution system, which is of special value when such antennas are used on board aircraft.

(2) The desired amplitude distribution can readily be implemented (by adjusting the amount of coupling between the slots and the waveguide).

A drawback of slotted waveguide arrays is the limited bandwidth above all because of the undesirable beam deviations with changes in frequency (in nonresonant antennas) and a low quality of input match.

Chapter Twelve

Radiating Apertures and Arrays

12.1. Basic Relationships

Linear radiating systems shape a unidirectional beam and provide for scanning in only one plane passing through the antenna axis. If a unidirectional beam is to be produced in the other, mutually

perpendicular plane as well, the radiating system must be placed on some surface of sufficiently large dimensions as compared with the wavelength. This arrangement results in a unidirectional antenna with a continuous or discrete distribution of electromagnetic sources within the chosen portion of the surface, or the *aperture*. By applying the principle of equivalence (see the Appendix), the shape of the aperture may be deformed in any desired manner. This chapter will be concerned with the most commonly used and readily analysable plane apertures. The aperture may be rectangular, circular, elliptical or any other in contour.

Let us agree to specify a radiating element of an aperture by its complex vector radiation pattern $\dot{\mathbf{F}}_{el}(\theta, \varphi)$ in its own, or local, coordinate system. The origin of this coordinate system is located inside the element (for example, at its radiation centre), and the z-axis is at right angles to the aperture plane xOy. The element radiation pattern $\dot{\mathbf{F}}_{el}(\theta, \varphi)$ will be deemed the same for all aperture elements, which is equivalent to postulating the same radiating current density distribution inside the elements. Apart from the radiation pattern of one radiator, the aperture is also characterized by the manner in which its elements are disposed (continuously or discretely) and by the amplitude and phase distribution of the complex amplitudes of element excitations. As follows from the product pattern (pattern multiplication) theorem (see Sec. 9.6), the radiation pattern of a plane aperture may be written as

$$\dot{\mathbf{f}}(\theta, \varphi) = \dot{\mathbf{F}}_{el}(\theta, \varphi)\, \dot{f}_{\Sigma}(\theta, \varphi) \tag{12.1}$$

where $\dot{f}_{\Sigma}(\theta, \varphi)$ is the complex pattern of an array of isotropic point sources located at the positions of the element centres, or simply the array factor.

For a discrete array of N radiators located at points (x_n, y_n) of a plane aperture (Fig. 12.1), the equation for the array factor can readily be derived from the general expression (9.22) as follows:

$$\dot{f}_{\Sigma} = \sum_{n=1}^{N} \dot{I}_n \exp(j\beta R_n \cos\alpha_n) \tag{12.2}$$

where the difference in path length to the observation point $P(\theta, \varphi)$ between the rays drawn from the origin of coordinates and from the point x_n, y_n is, in accordance with Eq. (7.4), given by

$$R_n \cos\alpha_n = \sin\theta\,(x_n \cos\varphi + y_n \sin\varphi) \tag{12.3}$$

It is assumed in Eqs. (12.3) and (12.2) that all radiators within the aperture are consecutively numbered from 1 through N and that

$$\dot{I}_n = I_n \exp(j\Phi_n)$$

is the complex amplitude of the excitation current for the nth element.

If an aperture is filled by radiators in a continuous fashion, the summation in Eq. (12.2) may be replaced by integration with res-

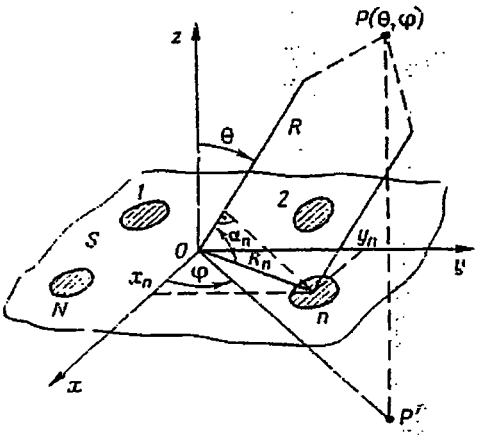


Fig. 12.1. To calculation of radiation from a planar aperture

pect to area, and the equation for the array factor will take the form

$$\dot{F}_\Sigma(\theta, \varphi) = \int_S \dot{I}(x, y) \exp[j\beta \sin\theta (x \cos\varphi + y \sin\varphi)]\, dx\, dy \qquad (12.4)$$

where S is the surface area of the aperture, and

$$\dot{I}(x, y) = I(x, y) \exp[j\Phi(x, y)]$$

is the excitation current distribution in amplitude and phase.

On introducing new angle variables (spatial frequencies)

$$\varkappa_1 = \beta \sin\theta \cos\varphi$$

and

$$\varkappa_2 = \beta \sin\theta \sin\varphi$$

we can reduce Eq. (12.4) to a bivariate Fourier transform of the excitation function

$$\dot{F}_\Sigma(\varkappa_1, \varkappa_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dot{I}(x, y) \exp j(\varkappa_1 x + \varkappa_2 y)\, dx\, dy \qquad (12.5)$$

such that the excitation current distribution $\dot{I}(x, y)$ is non-zero solely within the aperture surface S. Therefore, the *array factor* $\dot{F}_\Sigma(\varkappa_1, \varkappa_2)$ *is a band-limited, bivariate function.* The Fourier transform

of band-limited functions is widely used in telecommunications. Computation by Eq. (12.5) is carried out on a computer, using *fast Fourier transform* (FFT) algorithms. That is why it is convenient to treat the radiating systems of many types of unidirectional antennas as plane apertures varying in shape.

12.2. Directivity and Effective Area of a Planar In-Phase Aperture

Let there be a radiating system in the form of an opening of surface area $S \gg \lambda^2$, made in an infinite plane coincident with the $z = 0$ plane and separating the upper and lower half-spaces as shown in

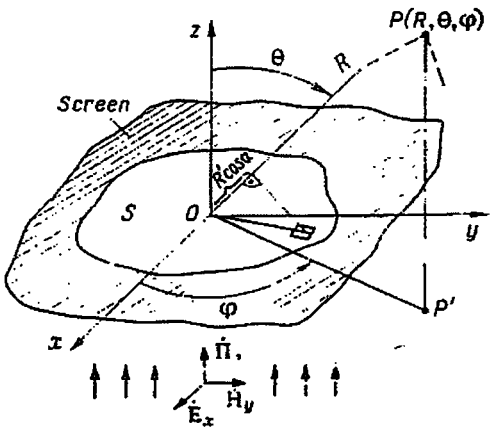


Fig. 12.2. Model of an in-phase aperture as a physical opening in a plane screen

Fig. 12.2. The opening is excited from the lower half-space by a plane electromagnetic wave propagated in the z-direction and made up of two electromagnetic field components $\dot{E}_x$ and $\dot{H}_y = \dot{E}_x/Z_m$, where $Z_m = (\mu_a/\varepsilon_a)^{1/2}$.

As follows from the equivalence principle [see Eqs. (A.10) and (A.11)], the electromagnetic field in the upper half-space can be treated as one produced by radiation from secondary sources distributed over some surface totally enclosing the true field sources. It is convenient to represent this surface by the $z = 0$ plane. It is a formidable task to determine the distribution of equivalent electric and magnetic currents in the $z = 0$ plane in rigorous terms, but physical intuition and knowledge of the applicable relationships from geometrical optics suggest that the total electromagnetic field in an opening of large dimensions closely resembles the field due to an undisturbed incident plane wave, whereas the currents on the up-

per shadow side of the infinite plane are negligible in comparison with the equivalent currents in the opening. Therefore, it is legitimate to replace the opening by a radiating aperture of surface area S and to neglect the radiation coming from all other portions of the $z = 0$ plane.

The aperture surface element $dS = dx dy$ is a *Huygens source* which produces, when $z > 0$ and R tends to infinity, an electromagnetic field $d\dot{\mathbf{E}}$ described by Eq. (A.16). On integrating the expression for $d\dot{\mathbf{E}}$ over the entire surface area of the aperture and taking into account the spatial path length difference given by Eq. (12.3), we obtain the following equation for the far-zone radiation field of the aperture:

$$\dot{\mathbf{E}} = (\mathbf{i}_\theta \cos\varphi - \mathbf{i}_\varphi \sin\varphi) \frac{I(1+\cos\theta)\exp(-j\beta R)}{2\lambda R} \int \dot{\mathbf{E}}_x(x, y) \times \exp[j\beta \sin\theta (x\cos\varphi + y \sin\varphi)]\, dx\, dy \quad (12.6)$$

where the aperture excitation $\dot{E}_x(x, y)$ is assumed of variable amplitude and of variable phase for generality.

When the aperture is excited in-phase, the radiation maximum is oriented along the z-axis where the path length difference is zero for all aperture elements. The magnitude of the vector $\mathbf{E}$ in the far-zone region in this direction is

$$E_{max} = \left| \int_S \dot{E}_x(x, y)\, dS \right| / \lambda R$$

In order to determine the directivity of an aperture it is essential to know also the total radiated power. The simplest way to find it is to evaluate the Poynting vector flow through the aperture surface area

$$P_\Sigma = \frac{1}{2} \int_S \dot{E}_x \dot{H}_y^* \, dS = \frac{1}{2Z_w} \int_S |\dot{E}_x(x, y)|^2 \, dS$$

where Z_m is the wave impedance of the medium. The computation yields an approximate result because the total field in an aperture is not exactly equal to that due to a plane electromagnetic wave.

Using the first definition of directivity in the form of Eq. (7.17), we obtain the following expression for the directivity of an aperture:

$$D = \frac{|E_{max}|^2 2\pi R^2}{Z_m P_\Sigma} = (4\pi/\lambda^2) \frac{\left| \int_S \dot{E}_x(x, y)\, dS \right|^2}{\int_S |\dot{E}_x(x, y)|^2 \, dS}$$

In the case of a constant-amplitude and constant-phase excitation distribution such that $\dot{E}_x = \text{const}$, the integrals are easy to evaluate, and the following result is obtained:

$$D_0 = 4\pi S/\lambda^2 \tag{12.7}$$

On comparing the above equation with Eq. (8.13), we may conclude that a *planar in-phase aperture with a constant-amplitude distribution has an effective area exactly equal to the surface area of the aperture.* By increasing the ratio S/λ^2, it is possible to raise the directivity of an in-phase aperture to very high values. For example, a square aperture with a side dimension of 10λ may have a directivity of 1250.

In the case of a variable-amplitude, variable-phase excitation distribution, the directivity of the aperture turns out to be smaller than its maximum value given by Eq. (12.7). This stems from Schwartz's inequality

$$\left| \int_S \dot{E}_x(x, y)\, dS \right|^2 \leqslant S \int_S |E_x(x, y)|^2\, dS$$

The reduction in directivity in the case of a non-ideal amplitude-phase distribution is usually stated in terms of what is known as the *aperture efficiency*, K_{ae}, given by

$$K_{ae} = S_{eff}/S = D/D_0 = \frac{\left| \int_S \dot{E}_x(x, y)\, dS \right|^2}{S \int_S |\dot{E}_x(x, y)|^2\, dS} \leqslant 1 \tag{12.8}$$

The value of K_{ae} does not depend on the shape of the element radiation pattern, therefore any excitation function $\dot{I}(x, y)$ may be used in Eq. (12.8) instead of $\dot{E}_x$.

12.3. The Array Factor of a Planar Aperture

Since the radiation pattern of a single aperture element in the half-space $z > 0$ is a nearly constant function [see the factor in front of the integral in Eq. (12.6)], the beamwidth and sidelobe level of an aperture are determined by the array factor defined in Eq. (12.4). To begin with, let us discuss a rectangular aperture of dimensions a by b (Fig. 12.3). The simplest case occurs when the current distribution may be pictured as the product of two factors, one depending solely on x and the other, solely on y, that is

$$\dot{I}(x, y) = \dot{I}(x)\, \dot{I}(y)$$

This is known as a *separable amplitude-phase distribution.* Then the array factor given by (12.4) can likewise be written as a product

$$f_\Sigma(\theta, \varphi) = \int_{-a/2}^{a/2} \dot{I}(x) \exp(j\beta x \sin\theta \cos\varphi)\, dx \times \int_{-b/2}^{b/2} \dot{I}(y) \exp(j\beta y \sin\theta \sin\varphi)\, dy \quad (12.9)$$

such that each factor is the same as the array factor of a linear antenna oriented along the x-axis in one case and along the y-axis in the other.

When all the elements of an aperture are excited equally and in phase, such that $\dot{I}(x) = \dot{I}(y) = \text{const}$, the integrals in (12.9) are easy to evaluate, and the array factor takes the form

$$F(\Psi_x, \Psi_y) = (\sin\Psi_x/\Psi_x)(\sin\Psi_y/\Psi_y) \quad (12.10)$$

where

$$\Psi_x = 0.5\,\beta a \sin\theta \cos\varphi$$

and

$$\Psi_y = 0.5\,\beta b \sin\theta \sin\varphi$$

A plot of the bivariate function defined by (12.10) is shown on the Ψ_x, Ψ_y plane, in Fig. 12.4 where, by virtue of symmetry, only one

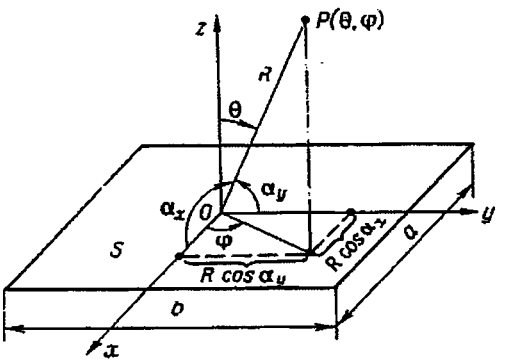


Fig. 12.3. Planar rectangular aperture

quadrant is depicted. The points mark the maxima of the main and side lobes. The numerals mark the levels of the lobe peaks. Around each side maximum, lobe sections at half-power levels are presented. The fine lines representing the null-radiation level are the boundaries of the individual lobes. The plot in Fig. 12.4 is invariant towards the aperture dimensions because they are accounted for by the angle

variables Ψ_x and Ψ_y. Similarly to an analysis of linear antennas, it is possible to isolate a *visible range* on the Ψ_x, Ψ_y plane. The boundaries of this range correspond to the directions of radiation in the aperture plane at $\theta = \pi/2$. The field in the lower half-space $\theta > \pi/2$ cannot be found by Eq. (12.6) or by Eq. (12.10), and this half-space is excluded from consideration. The constant θ lines on the Ψ_x, Ψ_y plane are ellipses such that

$$[\Psi_x/(0.5\beta a)]^2 + [\Psi_y/(0.5\beta b)]^2 = \sin^2\theta$$

and the boundary of the visible range at $\sin\theta = 1$ is specified in parametric form by

$$\Psi_{x,\,bound} = 0.5\beta a \cos\varphi$$

and

$$\Psi_{y,\,bound} = 0.5\beta b \sin\varphi$$

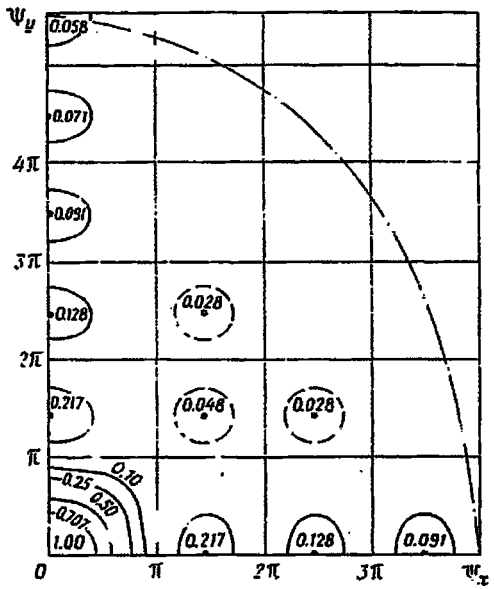


Fig. 12.4. Three-dimensional plot of the array factor of an ideal rectangular aperture

As an example, Fig. 12.4 shows the boundary of the visible range for $a = 4\lambda$ and $b = 5.5\lambda$. As the aperture increases in size, the visible range expands, and an ever greater number of side lobes fall within it. The highest side lobes occur in the main xOz and yOz planes. This finding remains valid for cases where the elements of an aperture are excited by currents varying in amplitude and in phase, as well. That is why analysis of the spatial array factor for a rectangular aperture is usually limited to a study of the planes where it is cut by the main xOz and yOz planes. In each of these planes the array factor is in effect the radiation pattern of a linear antenna similar to that examined in Chap. 11. The estimates obtained there for the beamwidth, the sidelobe level, the effect of the form of the excitation function on the shape of the array factor remain applicable to the main planes of a rectangular aperture. Notably, when the elements of an aperture are all excited by currents equal in amplitude and in phase, the half-power beamwidth is

$$\Delta\theta_x = 51^\circ\lambda/a$$

in the xOz plane, and

$$\Delta\theta_y = 51^\circ\lambda/b$$

in the yOz plane. On deriving expressions for a and b from the above relations and substituting them in the equation for directivity, Eq. (12.7), we obtain

$$D_0 = 4\pi ab/\lambda^2 = 32700/(\Delta\theta^\circ_x \Delta\theta^\circ_y)$$

which confirms the approximate equation (7.20).

The maximum directivity of a rectangular aperture may alternatively be written as the product of three factors

$$D_0 = 4\pi ab/\lambda^2 = \pi(2a/\lambda)\,(2b/\lambda) = \pi D_x D_y$$

where

$$D_x = 2a/\lambda$$

and

$$D_y = 2b/\lambda$$

are the directivities of ideal linear antennas of dimensions $a \gg \lambda$ by $b \gg \lambda$. The factor π may be treated as the equivalent directivity of one aperture element.

When the excitation of the aperture elements is unequal in amplitude but separable along the coordinates x and y, the directivity of a

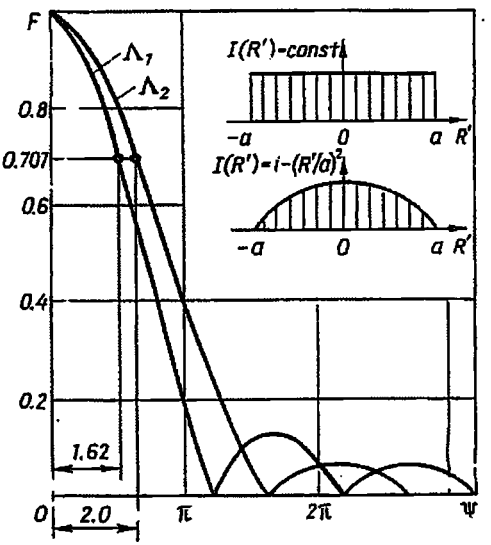


Fig. 12.5. Array factors of a circular in-phase aperture in the case of a constant-amplitude distribution Λ_1 and a tapered amplitude distribution Λ_2

rectangular aperture is impaired, and the resultant aperture efficiency can be estimated by the equation

$$K_{ae} = K_{ae,x} K_{ae,y}$$

where $K_{ae,x}$ and $K_{ae,y}$ are the aperture efficiencies of equivalent linear antennas, parallel to the x and y axes.

Let us estimate the main lobe efficiency and the scatter fraction of a rectangular aperture. It can be shown that the fraction of the total power radiated by the main lobe, that is, the main lobe efficien-

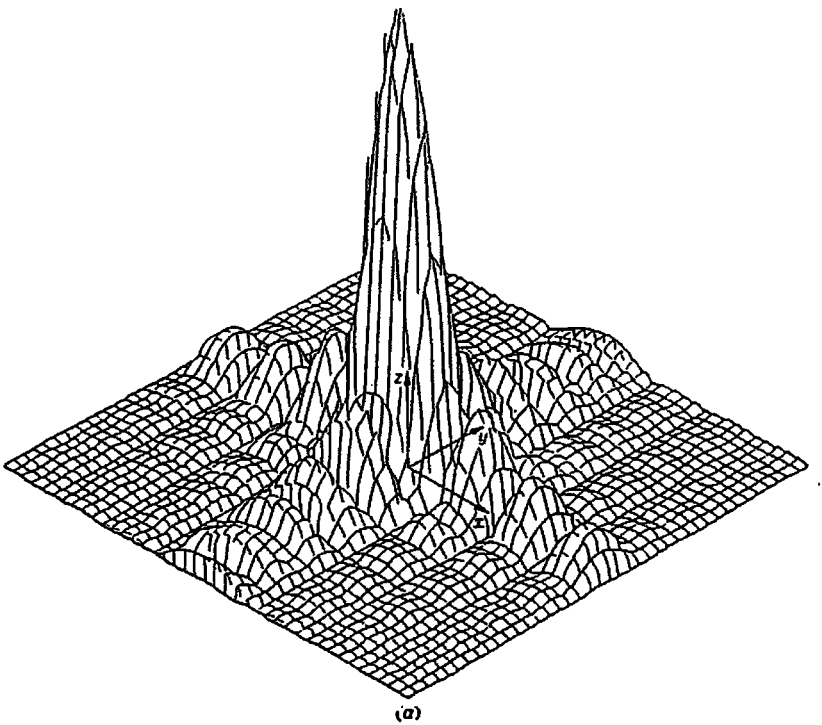


Fig. 12.6. Three-dimensional plots of the array factor for (*a*) a rectangular aperture and (*b*) a circular aperture

cy when a and b are substantially greater than a wavelength is approximately equal to the product of the main-lobe efficiencies of equivalent linear antennas parallel to the x and y axes,

$$(1 - \beta_s) \approx (1 - \beta_x)(1 - \beta_y)$$

where β_s is the scatter fraction of the rectangular aperture, whereas β_x and β_y are the scatter fractions of the equivalent linear antennas. Thus, the total scatter fraction of a rectangular aperture is

$$\beta_s = \beta_x + \beta_y - \beta_x\beta_y$$

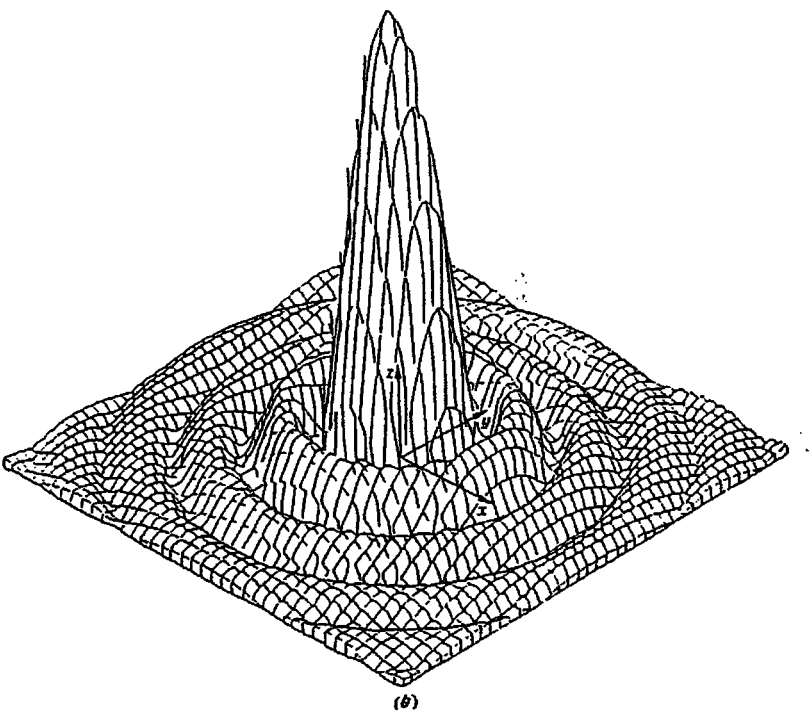
When the elements of an aperture are excited equally and in phase

$$\beta_x = \beta_y = 0.097$$

and

$$\beta_s \approx 0.185$$

The array factor of a circular aperture. When evaluating integrals of the form defined in (12.4), it is convenient to use spherical coordi-



(b)

nates and to write an element of surface as

$$dxdy = R'dR'd\varphi'$$

The path length difference in the exponent under the integral sign is

$$\sin\theta\,(x\cos\varphi + y\sin\varphi) = R'\sin\theta\cos(\varphi - \varphi')$$

and the equation for the array factor of a circular aperture of radius a takes the form

$$\dot{F}(\theta,\ \varphi) = \int_0^{2\pi}\int_0^{a} \dot{I}(R', \varphi')\exp[j\beta R'\sin\theta\cos(\varphi - \varphi')]\,R'\,dR'\,d\varphi' \quad (12.11)$$

Suppose that the excitation function is independent of the azimuthal angle

$$I(R') = (1 - \Delta) + \Delta[1 - (R'/a)^2]^n, \quad n = 1, 2, \ldots \quad (12.12)$$

where $1 - \Delta$ is the field level at the edge of the aperture, or what we will call the pedestal. Then the integral in (12.11) can be expressed in terms of the so-called lambda-functions $\Lambda_n(\Psi)$ as follows:

$$F(\theta) = 2\pi a^2 [(1 - \Delta)\Lambda_1(\Psi) + \Delta\Lambda_{n+1}(\Psi)/(n + 1)] \quad (12.13)$$

where $\Psi = \beta a \sin\theta$ is an angle variable analogous to the angle variables Ψ_x and Ψ_y used in the analysis of rectangular apertures.

Lambda-functions are connected to Bessel functions $J_n(\Psi)$ in a very simple way:

$$\Lambda_n(\Psi) = \frac{n! J_n(\Psi)}{(\Psi/2)^n}$$

The functions $|\Lambda_1(\Psi)|$ and $|\Lambda_2(\Psi)|$, which are the array factors of a circular aperture when the amplitude distribution of the excitation is $I(R') = \text{const}$ and $I(R') = 1 - (R'/a)^2$, are shown in Fig. 12.5. Note that the array factor of a circular aperture with an amplitude distribution of the form defined in (12.12) possesses axial symmetry and is independent of the angle φ. Thus, the rectangular aperture and the circular aperture have fundamentally different 3-D plots of sidelobe radiation (Fig. 12.6).

The half-power beamwidth (between the 0.707 points) of a circular aperture with its elements all excited equally and in phase is given by

$$\Delta\theta = 59^\circ\lambda/2a$$

when the sidelobe field level is 0.132 or —17.6 dB. If the amplitude distribution tapers towards the aperture edges, such that $1 - (R'/a)^2$, the beamwidth is increased by a factor of 1.23, and the sidelobe level falls to —24.7 dB. At the same time the aperture efficiency drops to 0.75, and the scatter fraction β_s goes down to about 0.02 as against the original value of $\beta_s \approx 0.162$ under conditions of a constant-amplitude distribution.

12.4. The Equivalent Linear Radiator Method in the Analysis of a Planar Aperture

Consider a planar aperture of an arbitrary shape such that the distribution of the excitation current $I(x, y)$ is fixed in advance in both amplitude and phase. If all the elements of the aperture are excited in phase, the direction of the main maximum is normal to its plane. The array factor of the aperture is given by the integral in Eq. (12.4). The quantity $x_q \cos\varphi_0 + y_q \sin\varphi_0 = u(\varphi_0)$ included

in the exponent of the integrand may be treated as the distance in the aperture plane from the origin of coordinate to the projection of the integration point Q on the u-direction specified by the angle φ_0 (Fig. 12.7). Noting this fact and changing to Cartesian coordinates u, v, z turned through the angle φ_0, it is an easy matter to reduce Eq. (12.4) to the form of the array factor for an equivalent linear antenna of length $L_{eq} = u_{max} - u_{min}$:

$$\dot{f}(\theta, \varphi_0) = \int_{u_{min}}^{u_{max}} \dot{I}_{eq}(u) \exp(j\beta u \sin\theta)\, du \qquad (12.14)$$

where the equivalent amplitude-phase excitation distribution is given by

$$\dot{I}_{eq}(u) = \int_{v_1(u)}^{v_2(u)} \dot{I}(u, v)\, dv$$

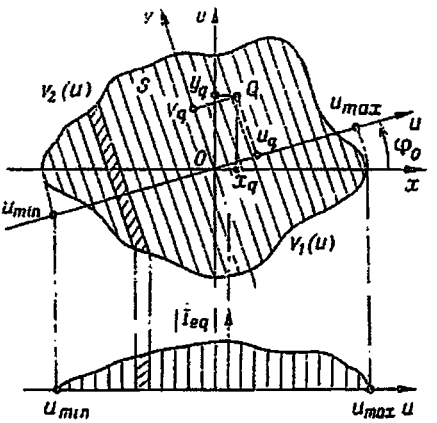


Fig. 12.7. To determination of the equivalent linear antenna concept

Here, $v_1(u)$ and $v_2(u)$ are the equations of the curves defining the lower and upper boundaries of the aperture. When the aperture elements are excited equally and in phase, the value of $\dot{I}_{eq}(u)$ is actually equal to the length of the chord passing through point Q parallel to the v-axis.

Example 12.1. For a circular aperture of radius a in which all the elements are excited equally and in phase, the equivalent amplitude and phase distribution in a linear antenna of length $L_{eq} = 2a$ has the form

$$I_{eq}(u) = \int_{-(a^2-u^2)^{1/2}}^{(a^2-u^2)^{1/2}} I_0\, dv = 2I_0 (a^2 - u^2)^{1/2} \quad \text{at} \quad |u| < a$$

This distribution tapers towards the aperture edges, and this explains why the main lobe broadens by as much as 16% and the sidelobe level goes down by 4.3 dB in a circular aperture as compared with a square aperture of a side dimension $2a$. The reduction in the sidelobe level and the broadening of the main beam in a circular aperture are not accompanied by a reduction in the aperture efficiency K_{ae} which is equal to unity when all the aperture elements are equally excited.

Example 12.2. For the diagonal plane of a square aperture of side dimension a in which all elements are excited equally and in

phase, the equivalent amplitude and phase excitation distribution in a linear antenna of length $L_{eq} = a\sqrt{2}$ is a linear function tapering towards the edges

$$I_{eq}(u) = I_0 a\sqrt{2}(1 - |u|) \quad \text{for} \quad |u| < a\sqrt{2}$$

On substituting this distribution in Eq. (12.14) and on integrating, we obtain the array factor

$$F(\theta, \pi/4) = \sin^2 \Psi/\Psi^2$$

where

$$\Psi = 0.707\, \beta a \sin\theta$$

for which the half-power beamwidth is

$$\Delta\theta \approx 52^\circ \lambda/a$$

and the sidelobe level is —26.5 dB which is a very low figure. The beamwidth in the diagonal plane of a square aperture is just by a mere 2% greater than the beamwidth in the principal planes. This is because the amplitude taper towards the edges of an equivalent linear antenna is made up for by an increase in its length by a factor of $\sqrt{2}$ as compared with the aperture dimension in the main plane.

Thus, resort to the concept of an equivalent linear antenna offers an opportunity to explain in simple and easy-to-grasp terms the effect that the shape of an aperture has on its directional characteristics in any plane passing through the z-axis.

12.5. The Effect of Random Excitation Errors on the Parameters of a Planar Aperture

In the presence of random errors, the distribution of excitation in amplitude and phase for an individual realization of an aperture has the form

$$\dot{I}(x, y) = \dot{I}_0(x, y)[1 + A(x, y)] \exp[j\Phi(x, y)] \qquad (12.15)$$

where $\dot{I}_0(x, y)$ = systematic part of the amplitude distribution
$A(x, y)$ = random function describing amplitude errors, of mean value zero and with a specified variance $\overline{A^2} \ll 1$
$\Phi(x, y)$ = random function describing phase errors, of mean value zero and with a specified variance $\overline{\Phi^2} \ll 1$

Apart from the variances, we deem known also the correlation radius (or length) ρ, defined as the distance between two points on the aperture, such that the random errors at those points are almost uncorrelated [see Eq. (11.22) and its commentary]. For each realization of the excitation function defined by (12.15) there is a correspon-

ding realization of the array factor defined by Eq. (12.4). On taking the ensemble average of realizations of random array factors, it is possible to establish the manner in which the aperture parameters depend on the total variance of the amplitude-phase distribution error

$$\alpha = (\overline{\Phi^2} + \overline{A^2}) \ll 1$$

and on the correlation radius ρ, similarly to how this was done for a linear radiating system in Sec. 11.5.

For a planar rectangular aperture of dimensions a by b and having all of its elements nominally excited equally and in phase, the average power radiation pattern is

$$\overline{|F(\theta, \varphi)|^2} \approx \exp(-\alpha)\{F_0^2(\theta, \varphi) + \alpha[4\pi^2\rho^2/D_0\lambda^2] \times \exp[-(\pi\rho \sin\theta/\lambda)^2]\}$$

where the first term on the right-hand side, $F_0(\theta, \varphi)$, is the normalized power radiation pattern of the aperture in the absence of random errors [see Eq. (12.10)], the second term is an excess sidelobe radiation slowly tapering down in all directions away from the main maximum, and $D_0 = 4\pi ab/\lambda^2$ is the directivity of the aperture in the absence of random errors.

The same form may be given to the average power radiation patterns of apertures other than rectangular in shape and also with their elements excited nonuniformly (the form of the excitation distribution function is accounted for indirectly in the value of D_0). Thus, what we have obtained for the average radiation pattern is analogous to the results stemming from the analysis of how random phase errors affect the performance of a linear antenna. The excess sidelobe term decreases with increasing aperture dimensions. There is only a difference in the quantitative assessment of the sidelobe level which, given the same values of α and ρ, is lower in the case of a planar array.

As before, the probability of multilobe occurrence (excessive side lobes) can be evaluated by resort to the plots of Fig. 11.12. Instead of Eq. (11.24), however, use should be made of a similar equation for a planar aperture

$$\overline{f_s^2} = f_{s0}^2 + \alpha(4\pi^2\rho^2/D_0\lambda^2)$$

where f_{s0}^2 is the sidelobe level in the absence of random errors. Also, the parameter of the modified Rayleigh distribution, Eq. (11.25), should be taken as

$$\sigma^2 = \alpha(2\pi^2\rho^2/D_0\lambda^2)$$

instead of $\alpha\pi^{1/2}\rho/2L$ for a linear antenna.

The directivity of a planar aperture in the presence of random excitation errors and with $\rho \geqslant \lambda$ can be found by an equation analogous

to Eq. (11.23), that is

$$D \approx D_0 \exp(-\alpha)$$

where $D_0 = 4\pi S_{eff}/\lambda^2$ is the nominal directivity in the absence of random errors.

As the correlation radius decreases in comparison with the wavelength, the excess sidelobe level in a planar aperture goes down in proportion to $(\rho/\lambda)^2$, that is, faster than it does in the case of a linear antenna, so that with the ratio ρ/λ tending to zero the directivity of a planar aperture comes out to be

$$D = D_0 (1 - 3\pi^2 \alpha \rho^2/4\lambda^2)$$

This equation should be used when $\rho/\lambda < 0.35$.

12.6. Beam Scanning in a Planar Aperture

The spatial position of the beam radiated by a planar aperture can be controlled by causing the phase of the excitation to vary in a linear fashion. Assume the excitation applied to a planar rectangular aperture of dimensions a by b to be a superposition of two current waves travelling one along the x-axis and the other along the y-axis

$$\dot{I}(x, y) = \dot{I}_0 \exp[-j\beta(\xi_x x + \xi_y y)], \quad \beta = 2\pi/\lambda$$

where $|\xi_x| < 1$ and $|\xi_y| < 1$ are the respective relative phase velocities. Actually, this superposition of two travelling waves is equivalent to one travelling wave propagated in the direction

$$\varphi_0 = \arctan(\xi_y/\xi_x)$$

in the aperture plane.

On integrating, the normalized array factor defined by (12.4) takes the form

$$F(\Psi_x, \Psi_y) = \left| \frac{\sin \Psi_x}{\Psi_x} \frac{\sin \Psi_y}{\Psi_y} \right| \qquad (12.16)$$

where

$$\Psi_x = 0.5\beta a (\sin\theta \cos\varphi - \xi_x)$$

and

$$\Psi_y = 0.5\beta b (\sin\theta \sin\varphi - \xi_y)$$

The direction for the principal maximum of radiation is defined such that $\Psi_x = 0$ and $\Psi_y = 0$. This leads to what are called the *phasing equations*

$$\xi_x = \sin\theta_0 \cos\varphi_0$$

and

$$\xi_y = \sin\theta_0 \sin\varphi_0$$

These equations offer an opportunity to find the relative phase velocities that would cause the peak of the radiation pattern to be orien-

ted in a specified direction θ_0, φ_0, and they hold for any shape of the aperture and for any amplitude distribution.

The expressions $\sin\theta\cos\varphi = \cos\alpha_x$ and $\sin\theta\sin\varphi = \cos\alpha_y$ are the direction cosines of the angles α_x and α_y that the direction of observation specified by the angles θ and φ makes with the x- and y-axes of the aperture, respectively (see Fig. 12.3). Let us examine a

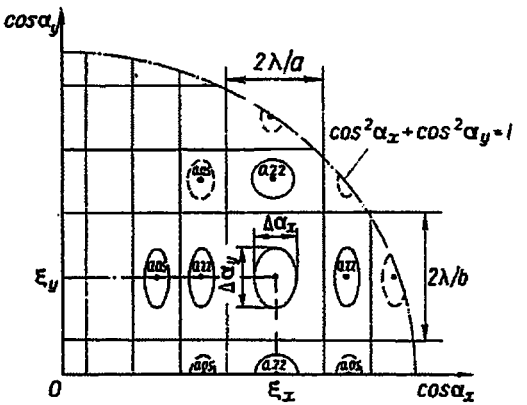


Fig. 12.8. Three-dimensional plot of the array factor of a rectangular aperture with a linear phase variation

three-dimensional plot of the array factor defined by Eq. (12.16) and constructed on the plane of the direction cosines (Fig. 12.8). The circle on which

$$\cos^2\alpha_x + \cos^2\alpha_y = 1$$

bounds the *visible range*. The peak of the radiation pattern is oriented towards the point whose coordinates are ξ_x, ξ_y. The section through the main lobe between half-power points can, to a good degree of approximation, be drawn as an ellipse with major axes

$$\Delta\alpha_x = 0.886\lambda/a$$

and

$$\Delta\alpha_y = 0.886\lambda/b$$

The dots in the plot indicate the positions of the side lobes. As the relative phase velocities vary, the lobes of the spatial array factor $F(\Psi_x, \Psi_y)$ are displaced on the direction cosine plane in a plane-parallel fashion so that all the lobes retain their original shape. This finding holds for apertures of any shape and with any amplitude-phase excitation distribution.

The direction cosine plane may be treated as a projection of a half-sphere of unity radius onto the equatorial plane coincident with the aperture plane (see Fig. 12.9). The position and shape of the half-

power section through the main lobe of the spatial array factor on a sphere of unity radius are determined by the intersection of this sphere with an elliptical cylinder which is parallel to the z-axis and has as its base the half-power section of the main lobe on the direction cosine plane (Fig. 12.9).

As the beam deviates from the normal to the aperture, the area bounded by the main-lobe section on the sphere of unity radius in-

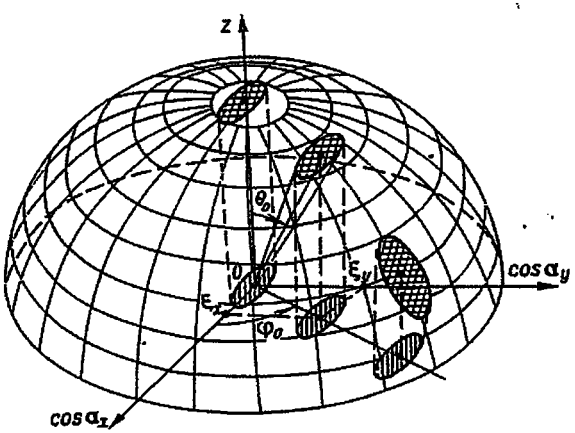


Fig. 12.9. Changes in the main lobe of the array factor of a rectangular aperture in scanning

creases in proportion to $1/\cos\theta_0$ as compared with an aperture excited by currents having a constant uniform phase. Because of this, the directivity of the aperture diminishes as

$$D = D_0 \cos\theta_0 = 4\pi S_{eff} \cos\theta_0/\lambda^2 \qquad (12.17)$$

where S_{eff} is the effective aperture with a constant uniform phase. Equation (12.17) holds on the condition that the direction of the main lobe does not approach the aperture plane nearer than two-three times the angular width of the main lobe.

The fundamental rule concerning the reduction of directivity in cosine scanning, Eq. (12.17), may be explained in a different way, using the equivalence principle. For this purpose, one should replace mentally the original aperture, *1*, with a linear phase variation, by an equivalent in-phase aperture, *2*, normal to the direction of the radiation maximum (Fig. 12.10). The area of this equivalent aperture decreases as $\cos\theta_0$, and this again leads to Eq. (12.17).

The reduction in directivity in scanning is accompanied by a broadening of the main lobe in the plane of beam deviation. If the beam is scanned in the principal planes of an equal-amplitude

rectangular aperture, the change in the beamwidth can be described by simple equations as follows:

$$\Delta\theta_x = 51°\lambda/(a \cos \theta_0)$$

and

$$\Delta\theta_y = 51°\lambda/(b \cos \theta_0)$$

For a square aperture similar equations hold with a good degree of accuracy for any scanning plane. However, when the beam is scanned in intermediate planes of a rectangular aperture (that is,

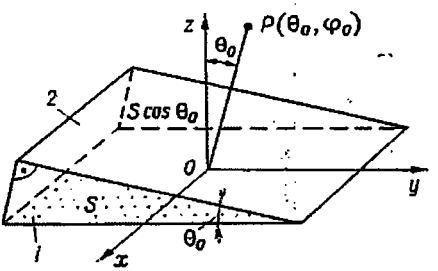


Fig. 12.10. Reduction in the effective area of an aperture in scanning

with $a \neq b$), the half-power contour on a sphere of unity radius takes the form of an ellipse with skewed major axes (see Fig. 12.9), and other equations should be used to obtain the beamwidth in the principal sections of this ellipse.

12.7. Flat Phased Arrays. Placement of Radiators Over the Aperture and the Conditions for the Absence of Grating Lobes

A linear phase variation on a planar aperture can be obtained by dividing the aperture into a great number of individual radiators such that the phase for each can be controlled independently of the others. The scanning system thus formed is referred to as a *phased array*.

An important issue in the design of a phased array is the choice of element placement in the aperture. Among the regular element arrangements, the most commonly used ones are those in which the elements are placed at the nodes of a rectangular or a triangular (hexagonal) lattice or grid (Fig. 12.11). With such an arrangement, the radiators are uniformly distributed over the aperture area or, which is the same, each radiator covers an equal fraction of the aperture area, either a rectangle of area given by

$$S_{el} = d_x d_y$$

in the case of a rectangular grid, or a hexagon of area given by

$$S_{el} = \sqrt{3}\, a^2/2$$

in the case of a triangular grid. (In the latter case, the area of the hexagon is equal to that of a parallelogram of side dimension a.)

Let us compare the two grids from the view-point of multilobe (grating lobe) occurrence in scanning. For this purpose, consider

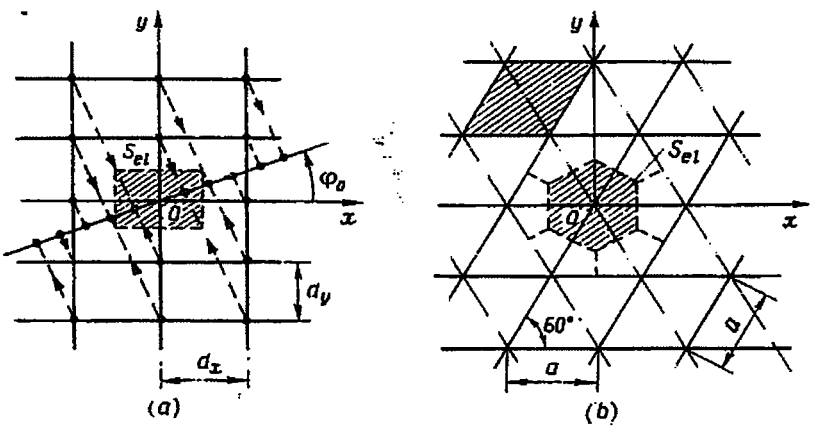


Fig. 12.11. Array radiators on (a) a rectangular grid and (b) a triangular grid

equivalent radiator rows formed by projections of the grid nodes on the direction of the aperture plane, specified by the angle φ_0 (Fig. 12.11a).

The rectangular grid. Here it will suffice to consider the unfavourable directions $\varphi_0 = 0$, $\varphi_0 = \pi/2$, and $\varphi_0 = \arctan(d_y/d_x)$ along which the equivalent rows have the largest constant spacings: d_x, d_y or $(d_x^2 + d_y^2)^{1/2}/2$. All the other directions give radiator rows with more closely spaced elements. The requirements that there should be no grating lobes in the visible range as the beam deviates through an angle θ_{max} from the normal to the aperture will, on the basis of Eq. (11.27), be satisfied if, with the number of elements N tending to infinity, the following constraints on the element spacing (grid pitch) are satisfied:

$$d_x < \lambda/(1 + \sin \theta_{x\,max})$$

and

$$d_y < \lambda/(1 + \sin \theta_{y\,max})$$

where $\theta_{x\,max}$ and $\theta_{y\,max}$ are the maximum angles of scan in the zOx and zOy planes, respectively. For example, with $\theta_{x\,max} = 45°$ it is essential that $d_x < 0.585\ \lambda$.

After d_x and d_y have been selected, one may check the allowable angle of beam deviation in an intermediate plane; it is always greater than either $\theta_{x\,max}$ or $\theta_{y\,max}$. With $\theta_{x\,max} = \theta_{y\,max}$, resort should be made to a square grid for which $d_x = d_y$ and in which the aperture surface area per grid element is given by

$$S_{el} \leqslant \lambda^2/(1 + \sin\theta_{max})^2$$

The triangular grid. The spacing between projections of its nodes on any direction in the aperture plane is less than or equal to $\sqrt{3}\,a/2$, and in choosing the element spacing the inequality

$$a < 2\lambda/[\sqrt{3}(1 + \sin\theta_{max})]$$

may be taken as a guide. For example, if the maximum angle through which the beam deviates from the normal is 45°, we find that $a \leqslant 0.68\lambda$. The aperture surface area per element for a triangular grid is

$$S_{el} = 2\lambda^2/[\sqrt{3}(1 + \sin\theta_{max})^2]$$

that is, by about 15% more than in the case of a square grid. This implies that the *total number of phased-array elements arranged in a triangular grid may be reduced by 13% as compared with their number in a square grid.* The increased element size in a triangular grid facilitates the placement of phase shifters near phased-array radiators.

The choice of element spacing for arrays using directional elements. As is noted in Sec. 11.6, grating lobes will be suppressed if they occur in the vicinity of a radiation null (or a nearly radiation null) of one element. This condition can be satisfied by giving the radiation pattern of an array element a special shape. Let the sector of scan (that is, the region enclosing the array beam) take up a range of angles such that $0 \leqslant \theta \leqslant \theta_{max}$ and $0 \leqslant \varphi \leqslant 2\pi$. Suppose that the three-dimensional radiation pattern of a radiator has the form

$$F^2_{el\,opt} = \begin{cases} \cos\theta & \text{for } \theta \leqslant \theta_{max}, \quad 0 \leqslant \varphi \leqslant 2\pi \\ 0 & \text{for } \theta > \theta_{max} \end{cases} \tag{12.18}$$

For this type of radiation pattern with an optimal suppression of grating lobes, the directivity is given by

$$D_{el\,opt} = 4\pi \Big/ \int_0^{\theta_{max}} \int_0^{2\pi} \cos\theta \sin\theta \, d\theta \, d\varphi = 4/\sin^2\theta_{max}$$

so that the effective area of one radiator is

$$S_{el\,opt} = D_{el\,opt}\,\lambda^2/4\pi = \lambda^2/(\pi \sin^2\theta_{max})$$

On dividing the aperture surface area equally among the optimal elements, we obtain their number over the entire array of a given

area S:

$$N_{opt} = S/S_{el\ opt} = \pi S\lambda^{-2} \sin^2 \theta_{max}$$

Let us compare the aperture area per element having an optimal radiation pattern with S_{el}, or the surface area per isotropic element, in a phased array arranged on a triangular grid, provided any grating lobes are suppressed. The ratio of these areas is

$$\sqrt{3}\,(1 + \sin \theta_{max})^2/\,(2\pi \sin^2 \theta_{max})$$

and it may be treated as the saving in the number of elements. Calculations show that a noticeable saving (by a factor of 2.5 and more) in the number of phased-array elements, on the condition that use is made of optimal directional radiators, can only be achieved with narrow sectors of scan, equal to or smaller than 30° at most. The design of radiators which have an optimal radiation pattern coming very nearly to that defined by Eq. (12.18) is a challenging engineering task. In practice, preference is given to a somewhat greater number of radiators. Apart from the fact that the phased-array radiators can then be made simpler in design, there is an additional improvement in reliability.

In practical phased arrays operating over small scan angles (less than 20° or 30°) the number of independently phased elements can be reduced in a simpler way by combining several simple radiators in in-phase groups each controlled by one phase-shifter. Each in-phase group has a narrower directivity pattern than one radiator, and this serves to reduce the level of grating lobes.

Unequal spacing of elements in flat phased arrays. It has been found that grating lobes can practically completely be suppressed in flat phased arrays if they use irregularly spaced radiators of low directional properties in a number substantially smaller than is the case with regularly spaced flat phased arrays. This approach is akin to the unequal spacing of elements in linear arrays. Most often, irregular placement is effected in some random manner. For example, the array aperture is first divided into a rectangular grid with $N \times N$ nodes. Then, M radiators are placed at some nodes, likewise chosen at random, so that every radiator has the same probability of $M/N^2 < 1$ that it will occur at any node. If the node spacing in the original grid is not over a half-wavelength, the average power radiation pattern will be the sum of the array factor of a flat regularly spaced array and a nearly uniform random 'background' due to the sidelobe radiation

$$\overline{F^2(\theta, \varphi)} \approx F_0^2(\theta, \varphi) + \frac{N^2 - M}{M(N^2 - 1)}\,[1 - F_0^2(\theta, \varphi)]$$

where $F_0^2(\theta, \varphi)$ is the radiation pattern of a closely spaced (or densely filled) regular array of $N \times N$ elements.

When N^2 is substantially greater than M, the average 'background' sidelobe radiation will be around $1/M$ and will be independent of N^2, that is, of the aperture dimensions. No grating lobes can occur because in any plane of scan the equivalent linear radiating system is an unequally spaced array containing all the M radiators. Thus, statistical analysis shows that in an ensemble of flat unequally spaced random arrays with the same number M of elements, the overwhelming majority of radiation patterns will be close to the average one, which means that antennas with a very poor and a very good radiation pattern have a very low probability of occurrence. Therefore, poor designs (for example, those in which the elements congest some area of the aperture) can readily be discarded by screening a limited number of random realizations.

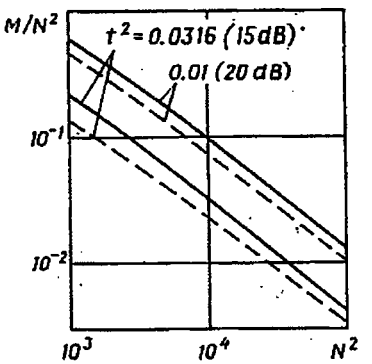


Fig. 12.12. To determining the number of elements for an unequally spaced planar array

When choosing the allowable number M of elements, the primary consideration should be the absence of side lobes that might exceed some predetermined level. Figure 12.12 shows plots which relate the allowable reduction in the number of elements, M/N^2, as a function of N^2 for two fixed values of sidelobe radiation t^2. The full curves apply to the values of M/N^2 at which the predetermined level is exceeded over the entire visible range by one random spike at most, and the dashed curves apply to the values of M/N^2 at which the specified level is exceeded by not more than 10 spikes. The plots hold if the node spacing in the original grid is a half-wavelength. Among other things, it follows from the above plots that if a $100 \times 100\lambda$ array, that is, one with $N^2 = 40\,000$, contains as few as 400 elements, the —15 dB level will be exceeded by not more than 10 lobes.

Let us define the aperture efficiency of a flat unequally spaced array as the ratio between its directivity and that of a flat aperture having the same area and energized with equal-amplitude and in-phase currents. The directivity of the aperture will be

$$D_0 = 4\pi S/\lambda^2 = 4\pi N^2 (0.5\lambda^2)/\lambda^2 = \pi N^2$$

The directivity of a widely spaced array is equal to the product of the directivity of one radiator by the number of elements

$$D = MD_{el}$$

because mutual coupling between widely spaced elements has almost no effect on the radiated power. For elements of low directional

properties (such as half-wave dipoles over a screen, half-wave slots, and the like), D_{el} is 3 or 4, and we may conclude that the area efficiency for a widely spaced array is approximately equal to the ratio M/N^2. Thus, so long as M is substantially smaller than N^2, unequally spaced arrays have a low aperture efficiency despite the fact that the main lobe of the radiation pattern is exactly the same in shape as the main lobe of a continuous aperture. The reduction in the aperture efficiency is due to the fact that the radiated power is scattered into a nearly uniform background of sidelobe radiation (see Sec. 7.5). Because of the low aperture efficiency, widely spaced arrays are seldom used in radars and are mostly employed in radio astronomy where the point that counts most of all is resolution whereas the low aperture efficiency can be made up for by an increase in the time of source observation.

12.8. Relationship Between the Radiation Pattern of an Array and Input Mismatch of Phased Array Elements in Scanning

In a phased array containing a small number of elements, the effect of mutual coupling between the elements on the array gain and the input impedance of an individual radiator in scanning can be analysed, using the 'element-by-element' approach examined in Sec. 11.8. If, on the other hand, a phase array has a large aperture with as many as several hundreds or even thousands of regularly spaced elements, it is convenient to use an approximate approach in which the average element characteristics are used. Then the total radiation pattern of the array will be defined by (12.1), *the function* $\dot{F}_{el}(\theta, \varphi)$ *being the radiation pattern of one element in the presence of other elements whose inputs are terminated in passive loads.* The resistance of these loads should be equal to the internal resistance of the generators (or receiving units) connected to each array element. In the circumstances, the excitation applied to only one element will, by virtue of mutual coupling, spread out to adjacent elements, thus giving rise to radiating currents in them. Also, some power will be lost in the passive loads. In consequence, the element radiation pattern $\dot{F}_{el}(\theta, \varphi)$ is in effect produced by a superposition of the fields due to the surrounding elements on that of a given element. The element efficiency with such a radiation pattern is $\eta_{el} < 1$ because some power is drawn by the loads of the adjacent elements. It follows, then, that the element radiation pattern $\dot{F}_{el}(\theta, \varphi)$ is connected to the element gain by a relation of the form

$$G_{el}(\theta, \varphi) = D_{el}\eta_{el}(1 - |\dot{\rho}_{el}|^2) F_{el}^2(\theta, \varphi)$$

where D_{el} is the element directivity as found from $F_{el}^2(\theta, \varphi)$, and $\dot{\rho}$ is the reflection coefficient at the input of the only driven element in the entire phased array.

The array factor $f_\Sigma(\theta, \varphi)$ appearing in (12.1), may, in the absence of grating lobes within the sector of scan, be approximately taken equal to the array factor of a continuous aperture. All estimates of beamwidth in scanning and of the aperture efficiency deduced on the assumption of random and systematic phase errors and unequal-amplitude excitation remain valid. The overall gain of a phased array may be written as

$$G(\theta_0, \varphi_0) = NG_{el}(\theta_0, \varphi_0) K_{ae,\Sigma} \qquad (12.19)$$

where (θ_0, φ_0) defines the direction of scan and $K_{ae,\Sigma}$ is the resultant value taking into account unequal-amplitude excitation, phase errors, and other factors. The effect of mutual coupling between the elements on the gain has been accounted for in the function $G_{el}(\theta, \varphi)$.

Consider the operation of a phased array when all of its elements are energized with constant-amplitude currents and are ideally phased in the direction defined by (θ_0, φ_0). Then in Eq. (12.19), $K_{ae,\Sigma} = 1$. Also, suppose that, while the element spacing is held constant, the aperture area is increased without bound and N tends to infinity. All elements will then be under identical conditions, and the manner in which the input reflection coefficient of each of them depends on the direction of scan θ_0, φ_0 will be described by the same function $\dot{\rho}(\theta, \varphi)$. The contribution of each element to the overall gain of a phased array may be written as the product of three factors, namely

(1) $4\pi S_{el}/\lambda^2$, which is the directivity of the aperture area per element.

(2) $\cos\theta_0$, which describes how the directivity decreases as the phased-array beam tilts from the normal to the aperture plane.

(3) $1 - |\dot{\rho}(\theta_0, \varphi_0)|^2$, which describes the decrease in the radiated power because of mismatch between radiator inputs in scanning.

Taking into account Eq. (12.19) and assuming that no grating lobes are formed, we may write

$$G_{el}(\theta, \varphi) = (4\pi S_{el}/\lambda^2) \cos\theta [1 - |\dot{\rho}(\theta, \varphi)|^2] \qquad (12.20)$$

The above equation is among the key ones in the theory of phased arrays. It establishes the connection between two modes:

(1) Only one radiator of an infinite array is energized, whereas all other elements are terminated in passive loads. For this mode, the radiation pattern is $G_{el}(\theta, \varphi)$, that is, one plotted in terms of gain.

(2) All elements of an infinite array are excited with equal-amplitude currents and with a linear phase variation in the direction

defined by θ, φ. For this mode, one uses the *dynamic input reflection coefficient* in the form of the function $\dot{\rho}$ (θ, φ).

It follows from Eq. (12.20) that the 3-D plot of the function $1 - |\dot{\rho}(\theta, \varphi)|^2$ replicates the directivity of one element in an infinite phased array accurate to cos θ. Therefore, there is no need to investigate the functions $G_{el}(\theta, \varphi)$ and $\dot{\rho}(\theta, \varphi)$ separately; it is sufficient to analyze any one of them. It is very convenient that the dynamic reflection coefficient function $\dot{\rho}(\theta, \varphi)$ should be defined in an infinite array—this offers an opportunity to use the theory of infinite periodic structures, well explored in electrodynamics, for purposes of phased array analysis and design.

To sum up, the design of large, regularly spaced flat phased arrays should use two electrodynamic models at the same time, namely:

(1) The flat continuous aperture model which takes into account the shape and dimensions of the aperture, the amplitude and phase distribution of excitation, and amplitude and phase errors.

(2) The infinite regularly spaced array model in which one defines the parameters of one radiator, notably the dependence of the dynamic reflection coefficient on the angle of scan and on frequency, and also selects and optimizes matching networks at the radiator inputs. Such an approach improves the accuracy of design in the case of large phased arrays because the number of isolated edge elements then constitutes only a small fraction of the total number of elements. Satisfactory results are obtained for phased arrays with a number of elements equal to 10 × 10 and more.

It has been found that the element radiation pattern $G_{el}(\theta, \varphi)$ in a poorly designed infinite array might decrease over the angle θ faster than the factor cos θ would suggest. In fact, even 'anomalous' nulls may appear at θ = 25 to 50°, although the conditions for multilobe occurrence have not yet been satisfied. Anomalous nulls are an indication of a complete mismatch between the array radiators because of the unfavourable interference of the waves that find their way at the radiator inputs via mutual coupling with other radiators (the phased array is 'blacked out' it might be said).

12.9. Transition-Zone Field of an Aperture. Focused Apertures

The electromagnetic field in the transition zone, that is, at a distance R from a radiating system less than $2D^2/\lambda$, where D is the maximum aperture diameter, is ordinarily calculated, using the *Fresnel approximation* [see Eqs. (7.10) and the relevant comments].

If we model an aperture as a physical opening in a plane (see Fig. 12.2), the field in the Fresnel approximation can be described by an equation similar to Eq. (12.6), with the only exception that the path length difference $\sin\theta\,(x\cos\varphi + y\sin\varphi)$ in the exponent of the integrand is augmented by a quadratic correction term dependent on the distance R: $-\rho^2\,(1 - \cos^2\alpha)/2R$, where

$$\rho = (x^2 + y^2)^{1/2}$$

is the radial distance from the centre of the aperture to the current point of integration on the aperture, and α is the angle at which the observation point and the integration point on the aperture are seen from the aperture centre. So long as the field observation angle (that is the angle between the direction of observation and the axis of the aperture) is less than or equal to 15°, the quantity $\cos^2\alpha \leqslant 0.067$ in the correction term may be neglected (this is what is known as the *Fresnel approximation for small angles*), and the remaining part of the correction term, $\rho^2/2R$, turns out to be independent of the observation angle.

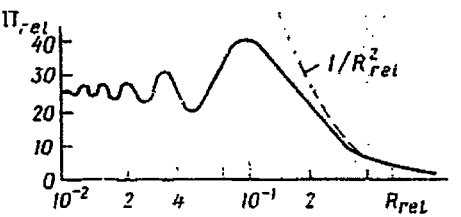


Fig. 12.13. Power flux density at the axis of a circular aperture in the case of amplitude taper $1 - (2\rho/D)^2_d$.

The Fresnel correction for small angles in Eq. (12.6) is equivalent in its effect to the occurrence, in the amplitude-phase aperture distribution, of quadratic phase errors, as measured at the aperture edge relative to its centre, $\Phi_{2eq} = -\beta D^2/8R$, where D is the largest diameter of the aperture and $\beta = 2\pi/\lambda$. Therefore, there is no need for a special analysis of the angular distribution of the aperture field in the transition zone—this distribution follows the shape of the array factor of the aperture in the presence of quadratic phase errors. On moving from the far-zone region into the transition zone of an aperture having a constant phase variation, the angular field distribution first shows 'null filling' and main-beam broadening. As the distance is reduced still more, the main lobe of the angular field distribution splits in two and partly merges with the side lobes. Thus, there is a complete analogy with the changes that take place in the radiation pattern shown in Fig. 11.9 as the quadratic phase error Φ_2 increases.

Characteristic changes take place in the magnitude of the Poynting vector on the z-axis of an in-phase aperture. The results found by calculation for a circular aperture with an amplitude taper of $1-4\rho^2/D^2$ are represented by the full curve in Fig. 12.13. Here the distance R is normalized to $R_{fzr} = 2D^2/\lambda$ and at this distance (that is, the boundary of the far-zone region) the magnitude of the Poyn-

ting vector is assumed to be equal to unity. The manner in which the Poynting vector customarily varies in the far-zone region, $1/R_{rel}^2 = (R/R_{fzr})^{-2}$, is shown by the dashed curve. This does not apply, however, in the transition zone. The Poynting vector oscillates on the z-axis and, closer to the aperture, it tends to a constant value determined by the amplitude distribution of the excitation over the aperture.

The oscillation of the Poynting vector on the aperture axis in the transition zone can readily be explained by the interference of the

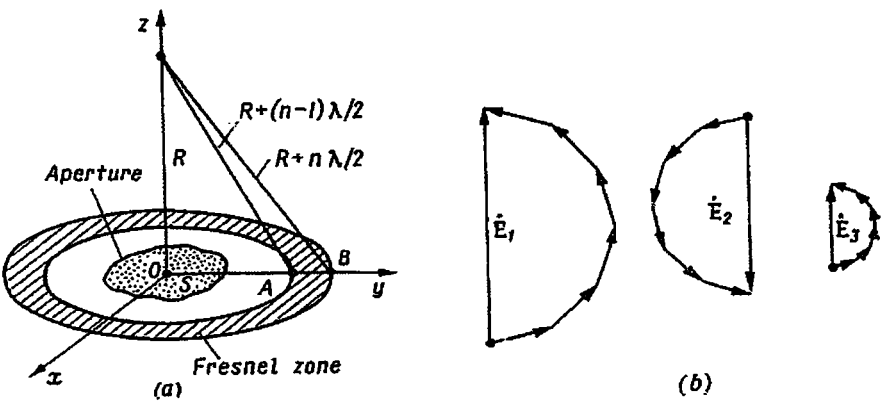


Fig. 12.14. Fresnel zones: *a*, construction of the *n*th zone; *b*, likely contributions of various zones to the total transition-zone field

waves that arrive at the observation point from different Fresnel zones on the aperture plane. A *Fresnel zone* is a circular region in the aperture plane xOy, selected so that radiation from its edge elements reaches the observation point at a common phase within π (180°). For an in-phase aperture, the difference in distance from the edges of a Fresnel zone (points A and B in Fig. 12.14*a*) to the observation point P should be equal to $\lambda/2$. The zones are reckoned from the projection of the observation point on the aperture plane, taking this projection as the zeroth zone. For the observation point located on the axis of an in-phase aperture within a distance R of its centre, the external radius of the nth Fresnel zone is given by the obvious relation

$$\rho_n^F = [(R + n\lambda/2)^2 - R^2]^{1/2} = [nR\lambda + (n\lambda/2)^2]^{1/2} \approx (nR\lambda)^{1/2}$$

The individual Fresnel zones turn out to be almost the same in surface area

$$S_n^F = \pi\,[(\rho_n^F)^2 - (\rho_{n-1}^F)^2] = \pi R\lambda + (2n - 1)\,\lambda^2/4 \approx \pi R\lambda$$

On dividing each Fresnel zone into a number of elementary circular sources and adding together at the observation point their respe-

ctive radiations with due account of their phases which are all different because of the difference in distance to point P, we can find the contributions $\dot{E}_1$, $\dot{E}_2$, . . . from the individual Fresnel zones to the total field at the observation point (Fig. 12.14*b*). It is seen that the vectors $\dot{E}_i$ due to adjacent Fresnel zones are opposite in direction while their amplitudes decrease with increasing zone number and also because the aperture excitation has an amplitude distribution which tapers down towards the edges as the distance to point P increases. The resultant field strength vector at the observation point is the sum of components opposite in direction

$$\dot{E} = \dot{E}_1 + \dot{E}_2 + \dot{E}_3 + \ldots$$

whose number is the same as that of Fresnel zones falling within the limits of the antenna aperture. For a circular in-phase aperture of diameter D, the number of Fresnel zones is given by the ratio between the surface area of the aperture and the surface area of one zone, that is,

$$n_F = \pi D^2/4\pi R\lambda$$

On expressing the distance to the observation point in relative units on a per-unit basis)

$$R_{rel} = R/(2D^2/\lambda)$$

we find that

$$n_F = 1/8\ R_{rel}$$

At observation points in the far-zone region, the aperture accounts for only a small fraction of the first Fresnel zone, while the remaining zones on the xOy plane are nonradiating and no interference takes place. As the distance to the observation point is reduced, the number of Fresnel zones on the aperture increases, and interference takes places between the waves radiated by the individual zones. In consequence, alternating maxima and minima appear in the axial distribution of the Poynting vector (see Fig. 12.13). An odd number of Fresnel zones on the aperture corresponds to maxima in the axial distribution of the magnitude of the Poynting vector, whereas an even number of Fresnel zones corresponds to minima in this distribution. The depth of the minima depends on the degree of amplitude taper towards the edges of the aperture. The deepest minima (almost as far as the zero level) occur in the case of equal-amplitude excitation.

The focused aperture. If we extend the amplitude-phase distribution of an in-phase aperture to include a progressive phase advance such that the phase lead as measured at the aperture edge is $\Phi_2 = \beta D^2/8R_F$, where $\beta = 2\pi/\lambda$, then the main part of the Fresnel correction for the transition zone, $-\beta\rho^2\,(1 - \cos^2\alpha)/2R$, will be

balanced out at an observation point $z = R_F$ on the aperture axis, and *the angular field distribution at this distance will be exactly the same as the radiation pattern of a constant phase aperture in the far-zone region.* (This statement holds true for $\cos \alpha \ll 1$, that is, for observation angles $\theta \leqslant 15°$ to 20°.) In the circumstances, it is customary to deem that the aperture is focused to a point $z = R_F$, and this point is called the *focus* of the aperture. For a constant-phase aperture, the focal point lies at infinity.

The characteristics of a focused aperture are used in cases where it is desirable to focus an antenna at a finite distance for the purpose of measuring the radiation pattern at a much closer range to the antenna than is possible for an antenna focused at infinity. In such cases, it is possible to reduce by a factor of several tens the distance at which the angular field distribution replicates the shape of the central part of the radiation pattern by merely introducing deliberate phase errors into the aperture of the antenna under test so that the phase lead is progressively increased towards the aperture edge. In the case of reflector or lens antennas, focusing to the desired point in the transition zone can be obtained by just shifting the feed from the focal point along the aperture axis away from the reflector or lens. Focused apertures are also used in radar to enhance resolution in the detection of targets lying in the transition zone of the aperture, in wireless power transmission over relatively short distances, in units which use concentrated electromagnetic energy for material processing, and in testing dielectrics and the plasma by 'radio-raying'. In multielement antenna systems, focusing is effected by ensuring suitable phase lags or leads at the inputs of the individual elements and also through the choice of a suitable shape for the surface on which the radiators are arranged.

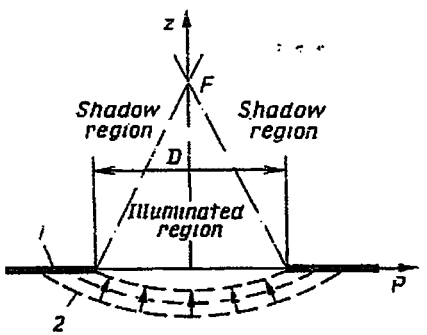


Fig. 12.15. Model of a focused aperture as a physical opening in a plane screen

From an electrodynamic point of view, a focused aperture is satisfactorily modelled by a physical opening of large electrical dimensions in a flat screen, *1*, excited by a converging spherical wave, *2* (Fig. 12.15). From the view-point of geometrical optics, all of the power radiated by a focused aperture is concentrated at the focal point, and the conical surface passing through the focus and the edges of the aperture provides a sharp boundary between illumination and darkness. Actually, such a boundary is non-existent, and electromagnetic radiation diffuses, in a way, from the illuminated

region into the surrounding space. In the vicinity of the focus, the region where the electromagnetic field is concentrated has finite dimensions, and the power density is thus limited.

The degree to which the electromagnetic field is concentrated by a focused aperture can be assessed, after B.G. Belyaev [17], in terms of the *boundary of the main power flow*. Here, the main power flow refers to that part of the radiated power which, in the case of an in-phase aperture, passes into the main lobe of the radiation pattern. (The ratio of the main power flow to the total radiated power is nothing but the efficiency of the main beam of the antenna, $1 - \beta_s$, where β_s is the scatter fraction which shows how much power is scattered into the side lobes.) Let the boundary of the main power flow be a surface enclosing the part of space inside which the main power flow is located. In order to obtain a universal main power flow distribution pattern applicable to apertures of an arbitrary diameter $D \gg \lambda$, it is useful to introduce normalized cylindrical coordinates $\rho^* = 2\rho/D$ and $z^* = z\,(4\Psi_0/\beta D^2)$, where $\beta = 2\pi/\lambda$, and Ψ_0 is the value of the generalized angle variable $\Psi = 0.5\,\beta D \sin\theta$ corresponding to the first null of radiation of an in-phase aperture*. With this choice of coordinates, the radius of the aperture is normalized to unity, and the angle of null radiation (that is, the half-width of the main lobe between adjacent nulls) of an in-phase aperture is equal to 45°, irrespective of the electrical dimensions of the aperture and the form of the amplitude distribution. The standard beginning of the far-zone region

$$z_{fzr} = 2D^2/\lambda$$

corresponds to the normalized coordinate $z^*_{fzr} \approx 5$. In the normalized coordinates ρ^*, z^*, the main power flow boundary for a circular aperture is approximately given by

$$\rho^*\,(z^*) = [0.9^{5/3}\,|\,1 - z^*/z^*_{Fd}|^{5/3} + (z^*)^{5/3}]^{0.6}$$

where z^*_F is the position of the focal point.

The position of the main power flow boundary defined by the above equation is invariant towards the form of the amplitude distribution—the form of the amplitude distribution solely affects the relative value of the main power flow, $1 - \beta_s$. The form of the main power flow boundaries in the normalized coordinates ρ^*, z^* is shown in Fig. 12.16. The unit segment on the ρ^* axis is the radius of the aperture, and the straight line drawn at an angle of 45° from the origin of coordinates is the direction of null radiation for an in-

* For a circular aperture energized with equal-amplitude and in-phase currents, $\Psi_0 = 3.83$. With an amplitude taper such that $[1 - (2\rho/D)^2]^\nu$, $\Psi_0 \approx = 1.25\,(\nu + 3)$, where $0.5 \leqslant \nu \leqslant 9$. The main beam efficiency is 0.838 for $\nu = 0$, 0.983 for $\nu = 1$, and 0.997 for $\nu = 2$.

phase aperture. The main power flow boundary of an in-phase aperture starts at a point near the edge of the aperture and gradually approaches an asymptotic value determined by the angle of null radiation. The portion of the main power flow at $z^* < 2.5$ is sometimes called, by an optical analogy, the 'searchlight beam' of the aperture. For a focused aperture, the main power flow boundary, starting at the same point on the aperture, tends to the line of the null radiation angle at a far steeper rate and touches it at $z = z_F^*$, which corresponds to the selected position of the focus. Past the focus, the main power flow boundary moves away from the line of the null radiation angle, which corresponds to a broadening of the main lobe of the angular field distribution at distances in excess of the focal length. The maximum compression of the main power flow of a focused aperture, corresponding to the minimum of the curve in Fig. 12.16, occurs markedly closer to the aperture than the focal point.

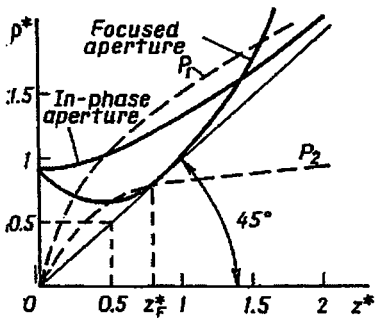


Fig. 12.16. Plots of the main power flow in a circular aperture

Main power flow boundary plots offer an opportunity to tackle many problems arising in power transmission from one aperture to another and also to determine accurately the size of the dangerous zone with a high level of electromagnetic radiation in front of the aperture of a transmitting antenna.

Example 12.3. Let there be an observation point P with coordinates ρ_P, z_P. This may, for example, be the edge of a receiving aperture arranged coaxially with a transmitting aperture. It is required to choose the dimensions of the transmitting aperture so that its main power flow boundary will pass through point P (this will ensure that the receiving aperture of a specified size will intercept the main power flow).

To begin with, construct an image of point P on the plane of normalized coordinates $\rho_P^* = 2\rho_P/D$, and $z_P^* = 4\Psi_0 z_P/\beta D^2$. On eliminating from these two equations the dimensions of the transmitting aperture, we obtain a relation which is in effect the equation of a parabola on the $\rho^* Oz^*$ plane

$$(\rho_P^*)^2 = z_P^* (\beta\rho_P^2/(\Psi_0 z_P)$$

Thus, *when the transmitting aperture has a variable diameter*, one observation point in real space is 'stretched' on the plane of normalized coordinates ρ^*, z^* into a parabola whose position depends on the coordinate ratio ρ_P^2/z_P, the wavelength, and the type of amplitude distribution (which is taken care of in the variable Ψ_0). As the

transmitting aperture increases in diameter, the image of the observation point moves along the parabola towards the origin of coordinates.

Figure 12.16 shows two parabolas, each corresponding to a distinct observation point. Parabola P_1 intersects the main power flow boundary of the unfocused aperture once in the transition zone and once in the far-zone region (outside the drawing). In consequence, the problem in such a case may have two solutions, each obtained with an in-phase aperture of a different size. It appears that preference should be given to a smaller aperture for which the observation point is located in the far-zone region. Parabola P_2 in Fig. 12.16 does not intersect the main power flow boundary of the in-phase aperture anywhere. Therefore, in order to concentrate the radiated power of the transmitting aperture within a circle of a specified radius, resort should be made to focusing. It is advantageous to select the focal point so that the transmitting aperture could have minimal dimensions.

Note that aperture focusing is also possible at distances z_F markedly shorter than the lower limit of applicability of the Fresnel approximation

$$R_{Fr} = D/4 + (D/\lambda)^{1/3}D/2$$

It should be remembered, however, that the minimal dimension of the focal spot is limited to about a half-wavelength (the wave limit of optics).

The range of problems that can be tackled by resort to main power flow boundary plots can be extended if one additionally considers the average parameters of amplitude and phase field distributions at various cross-sections of the main power flow [17].

Chapter Thirteen

Synthesis of Linear Antenna Systems

13.1. Synthesis of Radiating Systems

The objectives of antenna synthesis are to determine the shape and dimensions of a radiating system and to find its excitation distribution that would satisfy the specified requirements for the radiation pattern. Any one of three tasks may be set in antenna synthesis, as follows.

(1) To synthesize a radiation pattern having the specified shape.

(2) To synthesize a radiating system that would have either a maximal directivity or a maximal gain.

(3) To optimize the radiation pattern.

Let us dwell on each in more detail.

Synthesis of a radiation pattern having the specified shape. In this classical case, the shape of the required radiation pattern is specified by the well-known complex functions $\dot{G}(\theta, \varphi)$ and the objective is to find the excitation current distribution $\dot{I}(x, y, z)$ within a limited part of space that would yield a radiation pattern $\dot{F}(\theta, \varphi)$ which matches in some sense the specified function. Most commonly, the following two criteria of match between the functions $\dot{G}(\theta, \varphi)$ and $\dot{F}(\theta, \varphi)$ are used: the *minimum mean-square-error criterion* and the *minimum-deviation* (or *Chebyshev*) *criterion*. In the case of the minimum mean-square-error approach the synthesis error is estimated by the following integral:

$$\delta^2 = \int_{\Omega} |\dot{G}(\theta, \varphi) - \dot{F}(\theta, \varphi)|^2 \, d\Omega$$

and it is minimized by trying various forms of the excitation current distribution. The minimum mean-square-error criterion is used when the designer is interested in the directional properties of an antenna, averaged in an energy sense. A limitation of this approach is that the synthesized function may show appreciable local errors (spikes) even though the mean-square error is small (in mathematics, this is known as Gibbs' phenomenon).

In the case of the minimum-deviation (Chebyshev) approach the synthesis error is given by

$$\Delta = \max |\dot{G}(\theta, \varphi) - \dot{F}(\theta, \varphi)|$$

which is the maximum magnitude of the deviation of the synthesized radiation pattern from the specified function. As the value of Δ is minimized by trying various forms of the excitation current distribution, it is possible to secure a detailed match between $\dot{G}(\theta, \varphi)$ and $\dot{F}(\theta, \varphi)$, free from abrupt local spikes (overshoots) in the error. Unfortunately, the computation of Δ involves finding a maximum of the function of the radiation-pattern synthesis error, and this is far more difficult to do than to compute δ^2. For this reason, it is more difficult to solve the antenna synthesis problem, using the minimum-deviation (Chebyshev) criterion than the minimum mean-square-error criterion.

Synthesis of an antenna having a maximum directivity. In this case, the shape of the radiation pattern is not defined specifically, and various forms of the excitation current distribution are tried with a view to obtaining a maximum directivity in the desired direction. In addition to obtaining a maximum directivity, it may be required to obtain a null of radiation in some desired direction or to minimize the scatter fraction in a specified region of space.

Optimization of the shape of the radiation pattern. In this case, it is most often required to minimize the sidelobe level for a given width of the main beam or to obtain a minimal beamwidth for a specified sidelobe level.

In all the three cases of antenna synthesis the final step is to determine the excitation current distribution in the selected radiating system. Next comes the implementation of this solution by designing concrete elements for the radiating system and of the feed. Sometimes, this can be done, using dedicated computer-aided design programs compiled at the electrodynamic level. More often, however, one has to rely on empirical equations and to carry out experimental studies.

In our further discussion, we will limit ourselves to the synthesis of a linear radiating system, that is, an equivalent linear radiator of length $2l$, for which the array factor is given by the following equation [see Eq. (11.2) and the accompanying comments]:

$$\int_{-l}^{l} \dot{I}(z) \exp(j\varkappa z)\, dz = \dot{g}(\varkappa) \tag{13.1}$$

where

$$\varkappa = \beta \cos\theta$$

and the given right-hand side $\dot{g}(\varkappa)$ is defined solely in the visible range such that $-\beta \leqslant \varkappa \leqslant \beta$, that is, for $|\cos\theta| \leqslant 1$.

When a linear radiator is synthesized by invoking Eq. (13.1), it is actually assumed that the polarization characteristic of the antenna and the directional properties along the coordinate φ have already been secured through the proper choice of the radiating element and the only thing left to do is to select an applicable current distribution $\dot{I}(z)$. The unknown function $\dot{I}(z)$ is placed under the integral sign in Eq. (13.1), and so this equation is an *integral equation*. As is customary to classify in mathematics, this is an inhomogeneous integral Fredholm equation of the first kind. The function $\mathscr{K}(\varkappa, z) = \exp(j\varkappa z)$ is the *kernel* of this equation.

An unpleasant thing about Eq. (13.1) is the fact that however small changes in the function $\dot{g}(\varkappa)$ may correspond to however great differences in the excitation function $\dot{I}(z)$. For example, if for some function $\dot{I}_1(z)$ the corresponding radiation pattern is $\dot{f}_1(\varkappa)$, the function

$$\dot{I}_2(z) = I_1(z) + A \exp(jK_0 z/l)$$

will correspond to a radiation pattern given by

$$\dot{f}_2(\varkappa) = \dot{f}_1(\varkappa) + (2Al \sin\Psi)/\Psi$$

where

$$\Psi = \varkappa l + K_0$$

At the same time, for any large number A it is possible to select such a value of $K_0 \gg A$ that the functions $\dot{f}_1(\varkappa)$ and $\dot{f}_2(\varkappa)$ will be the same *within the visible range* despite the fact that the respective excitation distributions may show differences, however great.

The above example shows that the integral equation (13.1) belongs to what are called *incorrectly stated problems* likely to yield unstable solutions. In order to obtain a stable solution, some additional requirements stemming from the regularization principle introduced by A.N. Tikhonov of the USSR should be imposed. Notably, it may be required that the average excitation amplitudes and the average rates of change of the excitation along the antenna (the parameters A and K_0 in the example) be limited in comparison with the radiated power P_Σ

$$\left.\begin{aligned} \int_{-l}^{l} |\dot{I}(z)|^2 \, dz \leqslant M_1 P_\Sigma \\ \int_{-l}^{l} |\dot{I}'(z)|^2 \, dz \leqslant M P_\Sigma \end{aligned}\right\} \qquad (13.2)$$

where M_1 and M_2 are some predetermined constants. In addition to the conditions stated in (13.2), there are a number of other ways to avoid instability in the synthesized excitation distributions; this will be noted in our further discussion.

13.2. The Radiation Pattern as an Entire Function of a Finite Degree

Equation (13.1) is in effect the Fourier transform of the amplitude-phase excitation distribution function whose definition is completed with zero at $|z| > l$. For the function $\dot{I}(z)$ to have a Fourier transform, it should be square-integrable over an infinite interval and to have a finite number of maxima and minima over a finite interval. The first requirement means that

$$2\pi \int_{-l}^{l} |\dot{I}(z)|^2 \, dz = \int_{-\infty}^{\infty} |\dot{f}(\varkappa)|^2 \, d\varkappa = P < \infty \qquad (13.3)$$

where use is made of Parseval's theorem for Fourier integrals.

The second requirement is likewise obvious because a current distribution which would have an infinite number of maxima and minima along the length of an antenna cannot be realized. The

same constraints apply to the function $\dot{f}(\varkappa)$ for which we may write the following inverse Fourier transform:

$$\dot{I}(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \dot{f}(\varkappa) \exp(-jz\varkappa)\, d\varkappa \tag{13.4}$$

Let us take a closer look at the integral in (13.3) which is usually referred to as the total power of an antenna (accurate to within a constant). This integral may be re-cast as a sum

$$P = P_\Sigma + P_r = \int_{-\beta}^{\beta} |\dot{f}(\varkappa)|^2 d\varkappa + \left[\int_{-\infty}^{-\beta} + \int_{\beta}^{\infty}\right] |\dot{f}(\varkappa)|^2 d\varkappa$$

The first term is the integral over the visible range and gives the *active radiated power*. The second term contains the integrals over the entire range of imaginary angles and may arbitrarily be called the *reactive power*. This latter name is applicable because the function $\dot{f}(\varkappa)$ for $|\varkappa| > \beta$ describes the spectrum of slower-than-light non-radiating waves near the antenna, that is, it characterizes the reactive near-zone field of the radiating system.

The ratio of the total power of an antenna to the radiated power

$$\gamma = P/P_\Sigma = \frac{\int_{-\infty}^{\infty} |\dot{f}(\varkappa)|^2 d\varkappa}{\int_{-\beta}^{\beta} |\dot{f}(\varkappa)|^2 d\varkappa} = \frac{2\pi \int_{-l}^{l} |\dot{I}(z)|^2 dz}{P_\Sigma} \tag{13.5}$$

is the *reactivity ratio**. From comparison of Eqs. (13.5) and (13.2) it follows that the reactivity ratio is a measure of how the excitation function is incorrect, and its value should be limited when solving an antenna synthesis problem.

The Fourier transform of the function $\dot{f}(\varkappa)$ has a sharply limited spectrum because the excitation distribution should be non-zero only over the radiator length. This substantially narrows the class of functions representing the radiation pattern $\dot{f}(\varkappa)$. As the Paley-Wiener criterion states, the function $\dot{f}(\varkappa)$ integrable over the entire real axis and having a Fourier transform which is non-zero solely over the interval $[-l, l]$ represents on the complex $\varkappa$-plane an entire function of a finite degree l at most. In the theory of functions of a complex variable, the name 'entire' goes for functions which are

* Other authors call this quantity the supergain ratio. See, for example, "Antenna Theory" by R.E. Collin and F.J. Zucker, Vol. 1, p. 257. — *Translator's note.*

analytic in any limited domain. On the complex $\varkappa$-plane, an entire function has no singular points located at a finite distance from the origin of coordinates. The only singular point of an entire function is a point at infinity. As follows from the Paley-Wiener theorem, spectrum-limited functions are not just any entire functions, but only those which increase with increasing argument such that

$$| \dot{f}(\varkappa) | \leqslant \exp l | \varkappa |$$

The number l which characterizes the extent of the spectrum is called the degree or type of a given entire function. Examples of entire functions are $\sin \Psi/\Psi$, where

$$\Psi = \beta l (\cos \theta - \xi)$$

and also $\sin \varkappa l$, $\cos \varkappa l$, and various sums of these functions. The functions square-integrable over the entire real axis and satisfying the Paley-Wiener criterion are referred as the W_l class. The theory of antenna synthesis also uses the B_l class of entire functions which include functions that satisfy the Paley-Wiener criterion, but are not square-integrable over the real axis and are limited at all of its points. This class is essential in describing the array factor of uniform arrays.

Thus, the array factor of a linear radiator of length $2l$ is always an entire function of degree l at most. It has been found that entire functions of a finite degree can be used to approximate any continuous function $\dot{g}(\varkappa)$ on a finite interval of the $\varkappa$-axis to any level of accuracy using the minimum-deviation (or Chebyshev) match. This fundamental statement means that a linear antenna of length $2l$ can, at least in principle, serve to implement the array factor in the form of any given continuous function.

In this connection, let us estimate the derivatives in the array factor $\dot{f}(\varkappa)$ obtained by synthesis. On differentiating the array factor of a linear antenna, we get

$$| \dot{f}'(\varkappa) |^2 = \left| \int_{-l}^{l} jz\dot{I}(z) \exp(j\varkappa z)\, dz \right|^2$$

Application of the Cauchy-Bunjakowsky inequality gives

$$| \dot{f}'(\varkappa) |^2 \leqslant \int_{-l}^{l} | \dot{I}(z) |^2 dz \int_{-l}^{l} z^2 dz = \frac{2l^3}{3} \int_{-l}^{l} | \dot{I}(z) |^2 dz$$

Hence, in view of Eq. (13.5), we obtain

$$\max | \dot{f}'(\varkappa) | \leqslant (2l^3/3)^{1/2} (\gamma P_\Sigma / 2\pi)^{1/2}$$

Similarly, it is possible to derive inequalities in order to estimate any derivatives

$$\max | \dot{f}^{(n)}(\varkappa) | \leqslant [l^{2n+1}/(2n+1)]^{1/2} (\gamma P_\Sigma/\pi)^{1/2}$$

As follows from the above estimates, given a constant radiated power, an array factor with a steep slope can be produced either by increasing the radiator dimension $2l$ or, if the dimension is fixed in advance, by increasing the reactivity ratio γ, and this means resort to incorrect solutions of the integral equation (13.1). One consequence of such an approach is the occurrence of supergain which is discussed in Sec. 13.7.

13.3. Synthesis of a Linear Radiator by the Fourier Integral Method

Since the array factor of a linear radiator is the Fourier transform of the current distribution function, the following approach can be used in the synthesis of a linear radiator: specify the required radiation pattern $\dot{g}(\varkappa)$ and, by taking the inverse Fourier transform (13.4), find the current distribution

$$\dot{I}(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \dot{g}(\varkappa) \exp(-jz\varkappa)\, d\varkappa \qquad (13.6)$$

The required function is, however, known only within the visible range whereas the integral in (13.6) must be taken between infinite limits. Therefore, any one of two methods can be used to determine this function.

The first method. Choose the radiator length $2l$ and, using the techniques from the theory of functions of a complex variable, construct an analytic extension of the function $\dot{g}(\varkappa)$ from the visible range to the entire real $\varkappa$-axis. Take care that the extended function $\dot{g}(\varkappa)$ belongs to the W_l class of entire functions. Substitution of this function into (13.6) gives the only current distribution that is, in accordance with the Paley-Wiener theorem, non-zero solely within the radiator length and provides for an exact synthesis of the desired radiation pattern. The solution thus obtained may prove unstable because the analytic extension may lead to extremely high values of $\dot{g}(\varkappa)$ in the range of imaginary angles and this might in turn lead to an increased reactivity ratio. With this approach, one cannot regularize the solution by limiting the value of γ, whereas the manipulations involved in the construction of an analytic extension are complicated and might entail rounding-off errors in computations.

The second method. If one wishes to obtain a minimal reactivity ratio, it is reasonable to require from the outset that the function

$\dot{g}(\varkappa)$ should be zero outside the visible range, thus reducing the interval of integration. This gives a unique current distribution

$$\dot{I}_0(z) = \frac{1}{2\pi} \int_{-\beta}^{\beta} \dot{g}(\varkappa) \exp(-jz\varkappa)\, d\varkappa \tag{13.7}$$

where the zero in the subscript underscores the fact that the definition of $\dot{g}(\varkappa)$ is completed with zero so that it is extended into the range of imaginary angles. Because the function $\dot{g}(\varkappa)$ thus 'truncated' does not belong to the W_l class of entire functions, the amplitude-phase current distribution turns out to be nonzero over the entire $\varkappa$-axis, and a need arises to use, instead of it, some 'truncated' distribution $\dot{I}_z$ which would fit the radiation pattern that fails to match the specified function $\dot{g}(\varkappa)$. The distribution $\dot{I}(z)$ may be chosen so as to minimize the mean-square error

$$\delta^2 = \int_{-\infty}^{\infty} |\dot{g}(\varkappa) - \dot{f}(\varkappa)|^2\, d\varkappa = \int_{-\beta}^{\beta} |\dot{g}(\varkappa) - \dot{f}(\varkappa)|^2\, d\varkappa + \left[\int_{-\infty}^{-\beta} + \int_{\beta}^{\infty}\right] |\dot{f}(\varkappa)|^2\, d\varkappa \tag{13.8}$$

Applying Parseval's theorem to the above expression, we find

$$\delta^2 = 2\pi \int_{-l}^{l} |\dot{I}_0(z) - \dot{I}(z)|^2\, dz + 2\pi \left[\int_{-\infty}^{-l} + \int_{l}^{\infty}\right] |\dot{I}_0(z)|^2\, dz \tag{13.9}$$

For a given antenna length, the error is a minimum if

$$\dot{I}(z) = \begin{cases} \dot{I}_0(z) & \text{at} \quad |z| \leqslant l \\ 0 & \text{at} \quad |z| > l \end{cases}$$

Then the error δ^2 will solely be decided by the second term in (13.9), that is, by the excitation power on the discarded 'tails' which complete the radiator to an infinite straight line. The error can be reduced by extending the radiator, and this offers a convenient criterion for the choice of the length $2l$. Since in Eq. (13.9) the error is defined over the entire $\varkappa$-axis, the array factor found by the second method

$$\dot{f}_0(\varkappa) = \int_{-l}^{l} \dot{I}_0(z) \exp(jz\varkappa)\, d\varkappa \tag{13.10}$$

where the current $\dot{I}_0(z)$ is given by Eq. (13.7), will be small at $|\varkappa| > \beta$, which means that it will correspond to a reactivity ratio γ of nearly unity.

Thus, by the second method one obtains, instead of an exact solution, the best mean-square-error match to the required radiation pattern at the least possible reactivity ratio. This method of linear radiator synthesis is referred to as the *Fourier integral method*. It is extremely simple as regards computations and yields very stable current distributions readily realizable in practice. This last property has won the Fourier integral method wide popularity with antenna designers.

Comparison of the two methods for the synthesis of a linear radiator shows that they are in effect two mutually antithetic cases. In the first method, the synthesis error is zero but the reactivity ratio may be prohibitively large. In the second, the reactivity ratio is a minimum (this is known as overregularization) but the accuracy of synthesis is not too high (the minimum mean-square-error match). Standing in between them are an infinite number of solutions in which various trade-offs between the synthesis error and the reactivity ratio are possible.

Example 13.1. In a linear radiator of length $2l$ it is required to find the current distribution and the array factor that would match the Dirac delta-function as closely as possible:

$$\dot{g}(\varkappa) = \delta(\varkappa - \varkappa_0)$$
$$|\varkappa_0| \leqslant \beta = 2\pi/\lambda$$

at the least value of the reactivity ratio.

Using Eqs. (13.7) and (13.10), we find

$$\dot{I}_0(z) = (1/2\pi) \exp(-j\varkappa_0 z) \text{ at } |z| \leqslant l$$
$$f_0(\varkappa) = (jl/\pi) \sin \Psi/\Psi$$

where

$$\Psi = \beta l (\cos\theta - \varkappa_0/\beta)$$

That is, what we have obtained is the ideal linear radiator energized with equal-amplitude currents and in progressive phase that has been examined in detail in Chap. 11.

Using Eq. (13.5) to find the reactivity ratio for the found solution of the antenna synthesis problem gives

$$\gamma = D\lambda/4l$$

where D is the directivity of a linear radiator. In the normal (broadside) and tilted-beam radiation modes when the directivity only slightly exceeds $4l/\lambda$, the reactivity ratio is close to unity. In the axial radiation (end-fire) mode, the directivity is $8l/\lambda$ (at a relative phase velocity of $\xi = 1$), and the reactivity ratio rises to two.

13.4. Synthesis of Difference Radiation Patterns

A distinction of difference radiation patterns is the presence of two identical main lobes opposite in phase, adjacent to the boresight along which the radiation is zero. Such radiation patterns are emplo-

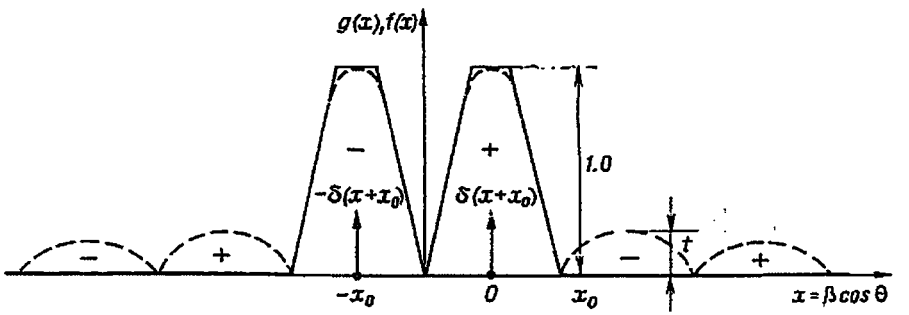


Fig. 13.1. Difference radiation patterns: ——— ideal; – – – – – – – – real

yed in the receiving channels of monopulse radars in order to generate an error signal proportional to the deviation of a target from the boresight.

The requirements that a difference radiation pattern (Fig. 13.1) should meet may be summed up as follows.

(1) The normalized radiation pattern should have the largest possible slope

$$S = \mathrm{d}F(\varkappa)/\mathrm{d}\varkappa$$

on the boresight.

(2) The directivity should be a maximum in the main lobes of the radiation pattern.

(3) The side lobes should be as small as practicable.

(4) The reactivity ratio should be a minimum. These requirements are, in a way, conflicting, and there is a need for some form of trade-off when implementing a specific difference radiation pattern.

Proceeding from the above requirements, let us specify the radiation pattern as the difference of two delta-functions (see Fig. 13.1) offset through $\pm\varkappa_0$ in opposite directions from the origin of coordinates

$$g_\delta(\varkappa) = [\delta(\varkappa - \varkappa_0) - \delta(\varkappa + \varkappa_0)]/\varkappa_0$$

On letting $\varkappa_0$ tend to zero, it is possible to obtain a maximum slope for the radiation pattern being synthesized at $\varkappa = 0$. Using Eqs. (13.7) and (13.10), find the amplitude-phase current distribu-

tion and the radiation pattern for a radiator of length $2l$

$$\dot{I}_0(z) = \frac{\exp(-j\varkappa_0 z) - \exp(j\varkappa_0 z)}{2\pi\varkappa_0} \equiv (\sin \varkappa_0 z)/\varkappa_0$$
$$\dot{f}_\delta(\varkappa, \varkappa_0) = \frac{\sin(\Psi - \delta\Psi)}{\delta\Psi(\Psi - \delta\Psi)} - \frac{\sin(\Psi + \delta\Psi)}{\delta\Psi(\Psi + \delta\Psi)} \quad (13.11)$$

where

$$\Psi = \beta l \cos\theta$$

and

$$\delta\Psi = \varkappa_0 l$$

In the limit of $\delta\Psi \to 0$, we obtain a difference radiation pattern which has a maximal slope and a minimal reactivity ratio

$$\dot{I}_\delta(z) = I_0 z$$
$$\dot{f}_\delta(\Psi, 0) = (\sin\Psi - \Psi\cos\Psi)/\Psi^2, \quad \Psi = \beta l\cos\theta \quad (13.12)$$

This radiation pattern and the associated current distribution are shown in Fig. 13.2 by the full lines. The function $f_\delta(\Psi, 0)$ has a

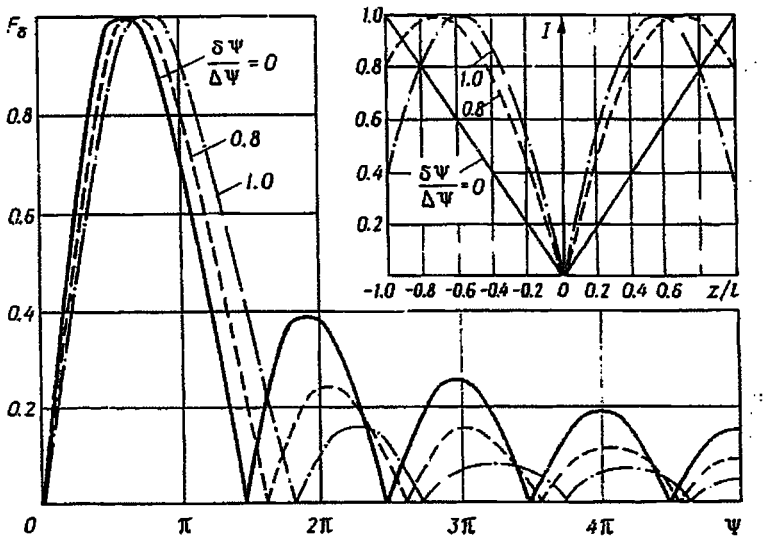


Fig. 13.2. To synthesis of difference radiation patterns

main peak at $\Psi_{max} = 2.081$ when it is equal to 0.435. The slope of $f_\delta(0, 0)$ at the origin of coordinates is one-third. In terms of true observation angles, the slope of the normalized difference radiation

pattern (in rad^{-1}) is

$$S_{\max} = |\, dF_\delta/d\theta \,|_{\theta=\pi/2} = 4.8l/\lambda$$

The sidelobe level of the difference radiation pattern defined by Eq. (13.12) is around 0.39, which is too much to be tolerated. The side lobes can be lowered by retaining non-zero values for $\delta\Psi$ in Eqs. (13.11). It is convenient to characterize $\delta\Psi$ in terms of a dimensionless quantity $\delta\Psi/\Delta\Psi$, where $\Delta\Psi = 2.78$ is the width of the main lobe of a reference sum radiation pattern, $\sin\Psi/\Psi$. The difference radiation patterns and amplitude-phase current distributions for several values of $\delta\Psi/\Delta\Psi$ are shown in Fig. 13.2 by the dashed lines. A plot of the directivity, slope and sidelobe level of a difference radiation pattern as functions of $\delta\Psi/\Delta\Psi$ is shown in Fig. 13.3. The maximum directivity in the main lobes of a difference radiation pattern equal to

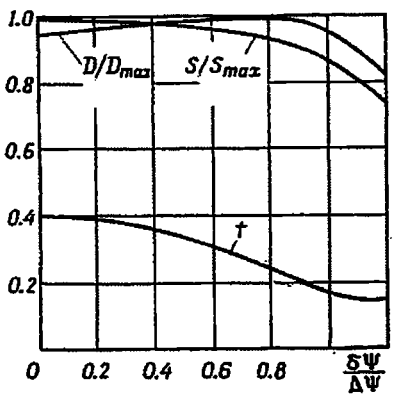


Fig. 13.3. Parameters of difference radiation patterns

$$D_{\max} = 2.46l/\lambda$$

is obtained at $\delta\Psi/\Delta\Psi \approx 0.8$, with the slope going down by a mere 7% and the sidelobe level going down to 0.25. As the ratio $\delta\Psi/\Delta\Psi$ keeps increasing, the slope decreases because the main lobes of the difference radiation pattern move farther apart, and the results then obtained are of no interest.

13.5. Synthesis of a Linear Radiator by the Partial Radiation Pattern Method

Let us write the current distribution in a linear radiator as a series expansion

$$\dot{I}(z) = \sum_{n=0}^{\infty} \dot{a}_n \dot{\varphi}_n(z) \quad \text{for} \quad |z| \leqslant l \tag{13.13}$$

of some known system of functions, $\dot{\varphi}_n(z)$. Let us substitute the above series in Eq. (13.10) for the array factor

$$\dot{f}(\varkappa) = \sum_{n=0}^{\infty} \dot{a}_n \int_{-l}^{l} \dot{\varphi}_n(z) \exp(j\varkappa z)\, dz = \sum_{n=0}^{\infty} \dot{a}_n \dot{f}_n(\varkappa) \tag{13.14}$$

The function

$$\dot{f}_n(\varkappa) = \int_{-l}^{l} \dot{\varphi}_n(z) \exp(j\varkappa z)\, dz, \quad \varkappa = \beta \cos\theta \tag{13.15}$$

which is dependent on the present number n represents the partial radiation pattern for which the excitation distribution has the form $\dot{\varphi}_n(z)$. Now, we can approximate the specified radiation pattern $\dot{g}(\varkappa)$ by the series expansion defined in (13.14), compute the required excitation coefficients $\dot{a}_n$, and finally find the required excitation distribution by Eq. (13.13). Taken together, these steps make up what is known as *synthesis by the partial radiation pattern method.*

It is easiest of all to implement the partial radiation pattern method in the case of the minimum mean-square-error match. Here, the system of functions $\dot{f}_n(\varkappa)$ should be some complete system of functions satisfying the condition of orthogonality

$$\int_{-\beta}^{\beta} \dot{f}_n(\varkappa)\, \dot{f}_p^*(\varkappa)\, d\varkappa = \begin{cases} N & \text{for } n = p \\ 0 & \text{for } n \neq p \end{cases}$$

Then the excitation coefficients $\dot{a}_n$ can be computed from the specified radiation pattern $\dot{g}(\varkappa)$ as generalized Fourier coefficients

$$\dot{a}_n = \frac{1}{N} \int_{-\beta}^{\beta} \dot{g}(\varkappa)\, \dot{f}_n^*(\varkappa)\, d\varkappa$$

Equation (13.15) shows that the functions $\dot{f}_n(\varkappa)$ are the Fourier transforms of the distributions $\dot{\varphi}_n(z)$ which are non-zero over the length of the radiator. Hence, the functions $\dot{f}_n(\varkappa)$ must be entire functions of degree l at most.

The theory of antenna synthesis widely uses systems of partial radiation patterns which are members of the Kotelnikov series and which offer a convenient means of representing entire functions of degree l at most over all of the $\varkappa$-axis

$$\dot{f}_n(\Psi) = \sin(\Psi - n\pi)/(\Psi - n\pi), \quad \Psi = \beta l \cos\theta$$

$$\dot{\varphi}_n(z) = \exp(jn\pi z/l), \quad |z| \leqslant l$$

Let us write the specified function $\dot{g}(\Psi)$ as a Kotelnikov series

$$\dot{g}(\Psi) = \sum_{-\infty}^{\infty} \dot{a}_n \sin(\Psi - n\pi)/(\Psi - n\pi)$$

A distinction of the Kotelnikov series is the fact that at the point

$\Psi_n = n\pi$ only one nth partial radiation pattern has a maximum of unity whereas all the other partial radiation patterns at that point are zero. Therefore, the unknown expansion coefficients $\dot{a}_n$ *are equally spaced samples of the given function*

$$\dot{a}_n = \dot{g}(n\pi)$$

As with the Fourier integral method, in order to determine all coefficients $\dot{a}_n$ the function $\dot{g}(\Psi)$ must be known along all of the

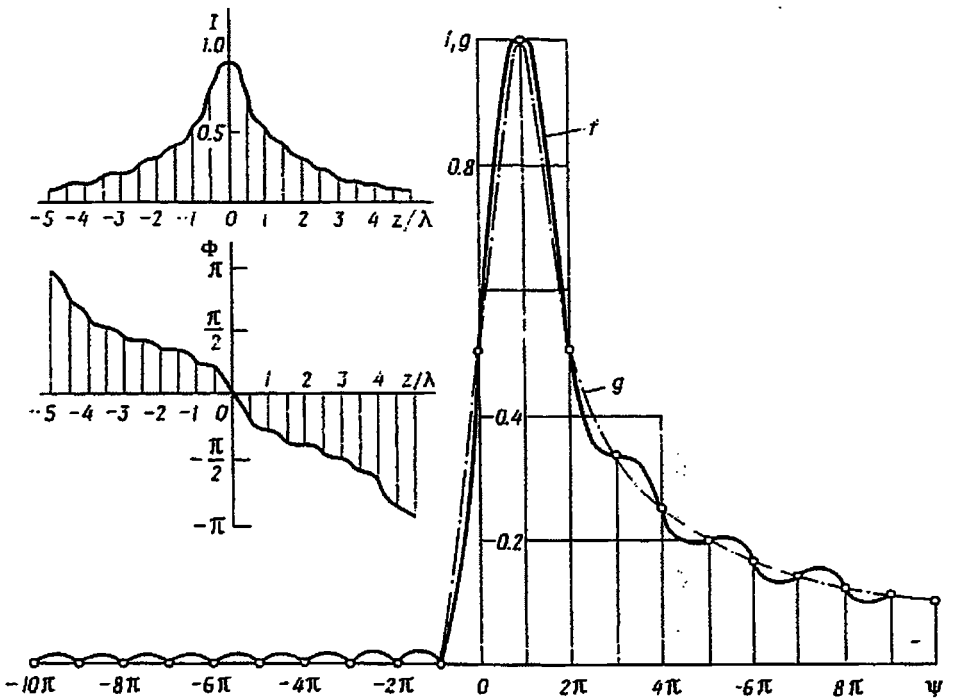


Fig. 13.4. Synthesis of a cosecant-squared beam

Ψ-axis. Again, any one of two approaches can be used to complete the definition of this function, namely:

(1) By constructing an analytic extension of $\dot{g}(\Psi)$ from the visible range to all of the Ψ-axis as an entire function of degree l at most—this would lead to an infinite number of coefficients $\dot{a}_n$ and a more accurately synthesized shape of the specified radiation pattern, but the reactivity ratio γ might then be prohibitively high (an incorrect solution).

(2) By completing the definition of $\dot{g}(\Psi)$ with zero at $|\Psi| > \beta l$, which would offer an opportunity to find only the first $2N + 1 =$

$4l/\lambda + 1$ coefficients of the Kotelnikov series, in which case the synthesized radiation pattern would have the form

$$\dot{f}(\Psi) = \sum_{-N}^{N} \dot{g}(n\pi) \sin(\Psi - n\Psi)/(\Psi - n\pi) \qquad (13.16)$$

where $N = \mathbf{E}(2l/\lambda)$, and $\mathbf{E}(x)$ is the integral part of the number x.

In the second approach, the required current distribution in amplitude and phase is given by a truncated Fourier series

$$\dot{I}(z) = \sum_{-N}^{N} \dot{g}(n\pi) \exp(j\pi z/l) \qquad (13.17)$$

The solution of the synthesis problem in the form of Eqs. (13.16) and (13.17) satisfies the requirement for a minimal reactivity ratio, but the synthesized radiation pattern $\dot{f}(\Psi)$ matches the specified function $\dot{g}(\Psi)$ only at sampling points, whereas at other values of the argument the specified and the synthesized radiation patterns are different.

Figure 13.4 illustrates, as an example, the synthesis of a cosecant-beam linear radiator often used in target tracking radars. The antenna is chosen to be 10λ long, which fact permits the use of 21 partial radiation patterns. The required radiation pattern $\dot{g}(\Psi)$ is shown by the dash-dot curve, and its samples which are amplitude coefficients in the expansions (13.16) and (13.17) are shown encircled. The synthesized radiation pattern precisely goes through the specified samples, but between the samples the synthesis is not precise.

13.6. Synthesis of Linear Arrays

According to Eq. (11.1), the array factor of a linear array has the form

$$\dot{f}_N(\theta) = \sum_{n=1}^{N} \dot{I}_n \exp(j\beta z_n \cos\theta)$$

where $\dot{I}_n$ = complex amplitude of the excitation current
z_n = radiator coordinate on the z-axis
N = number of array elements

Let us introduce the following notation:

$$\varkappa = \beta \cos\theta$$

and

$$\exp(j\beta z_n \cos\theta) = \exp(j\varkappa z_n) = \dot{F}_n(\varkappa)$$

The function $\dot{F}_n(\varkappa)$ may be treated as the normalized radiation pattern of one array element in a coordinate system common to all

elements. Hence, the radiation pattern of the array may be written as

$$\dot{f}_N(\varkappa) = \sum_{n=1}^{N} \dot{F}_n(\varkappa)\, \dot{I}_n = \langle \mathbf{F}(\varkappa)\, \mathbf{i} \rangle \tag{13.18}$$

where

$$\langle \mathbf{F}(\varkappa) = [\dot{F}_1(\varkappa),\ \dot{F}_2(\varkappa),\ \ldots,\ \dot{F}_N(\varkappa)]$$

is the row matrix of the element radiation patterns, and

$$\mathbf{i} \rangle = [\dot{I}_1,\ \dot{I}_2,\ \ldots,\ \dot{I}_N]_t$$

is the column matrix of the excitation coefficients.

In Eq. (13.18), the array factor is shown as a superposition of partial radiation patterns $\dot{F}_n(\varkappa)$, therefore the partial radiation pattern method is a natural choice for any array synthesis problems. An inconvenience of $\langle \mathbf{F}(\varkappa)$ is the fact that the elements of the matrix are non-orthogonal within the visible interval: $-\beta \leqslant \varkappa \leqslant \beta$.

To avoid this inconvenience, let us construct a new system of orthonormalized partial radiation patterns

$$\langle \mathbf{e}(\varkappa) = [\dot{e}_1(\varkappa),\ \dot{e}_2(\varkappa),\ \ldots,\ \dot{e}_N(\varkappa)]$$

which are linear combinations of all radiation patterns $\dot{F}_n(\varkappa)$

$$\dot{e}_n(\varkappa) = \dot{i}_{1n}\dot{F}_1\varkappa + \dot{i}_{2n}\dot{F}_2(\varkappa) + \ldots + \dot{i}_{Nn}\dot{F}_N(\varkappa)$$

where $\dot{i}_{np}$ are dimensionless coefficients. The condition of orthonormality means that

$$\frac{1}{2\beta}\int_{-\beta}^{\beta} \dot{e}_n \dot{e}_p^*\, d\varkappa = \begin{cases} 1 & \text{for } n = p \\ 0 & \text{for } n \neq p \end{cases} \tag{13.19}$$

Conversion from $\langle \mathbf{F}(\varkappa)$ to $\langle \mathbf{e}(\varkappa)$ can be written as a matrix equality of the form

$$\langle \mathbf{e}(\varkappa) = \langle \mathbf{F}(\varkappa)\, \mathbf{I} \tag{13.20}$$

where the columns of the square transformation matrix I are the expansion coefficients of each orthonormalized radiation pattern $\dot{e}_n(\varkappa)$

$$\mathbf{I} = \begin{Vmatrix} \dot{i}_{11} & \dot{i}_{12} & \ldots & \dot{i}_{1N} \\ \dot{i}_{21} & \dot{i}_{22} & \ldots & \dot{i}_{2N} \\ \ldots & \ldots & \ldots & \ldots \\ \dot{i}_{N1} & \dot{i}_{N2} & \ldots & \dot{i}_{NN} \end{Vmatrix}$$

The system of radiation patterns $\langle \mathbf{e}(\varkappa)$ forms an orthonormalized basis in the finite-dimensional space of functions which are all li-

kely factors of the array. Therefore, any radiation pattern may be written as an expansion of the form

$$\dot{f}_N(\varkappa) = \sum_{n=1}^{N} \dot{b}_n \dot{e}_n(\varkappa) = \langle \mathbf{e}(\varkappa)\ \mathbf{b} \rangle \tag{13.21}$$

where $\dot{b}_n$ are the expansion coefficients. It is also possible to go in the inverse direction from the array factor defined by (13.21) to the one defined by (13.18). To do this, substitute (13.20) into (13.21)

$$\dot{f}_N(\varkappa) = \langle \mathbf{F}(\varkappa)\ \mathbf{I}\mathbf{b} \rangle = \langle \mathbf{F}(\varkappa)\ \mathbf{i} \rangle$$

Hence,

$$\mathbf{i}\rangle = \mathbf{I}\mathbf{b}\rangle \tag{13.22}$$

Thus, given a known column matrix of expansion coefficients $\mathbf{b}\rangle$, the column matrix of the current distribution for the array is defined in terms of the same transformation matrix $\mathbf{I}$ whose rows are sets of expansion coefficients of each $\dot{I}_n$ over all $\dot{b}_n$

$$\dot{I}_n = \dot{i}_{n1}\dot{b}_1 + \dot{i}_{n2}\dot{b}_2 + \ldots + \dot{i}_{nN}\dot{b}_N$$

The solution of the array synthesis problem in the orthonormalized basis $\langle \mathbf{e}(\varkappa)$ presents no difficulties. To do this, let us specify the required radiation pattern as

$$\dot{g}(\varkappa) = \langle \mathbf{e}(\varkappa)\ \mathbf{b} \rangle = \sum_{n=1}^{N} \dot{b}_n \dot{e}_n(\varkappa) \tag{13.23}$$

On multiplying both sides of (13.23) by the complex conjugate function $\dot{e}_p^*(\varkappa)$ and integrating them with respect to $\varkappa$ between $-\beta$ and β, we obtain subject to the condition of orthogonality

$$\dot{b}_p = \int_{-\beta}^{\beta} \dot{g}(\varkappa)\, \dot{e}_p^*(\varkappa)\, d\varkappa$$

On collecting all $\dot{b}_p$ in one column and replacing the functions $\dot{e}_p^*(\varkappa)$ by linear combinations of functions $\dot{F}_n^*(\varkappa)$ in accordance with Eq. (13.20), we get

$$\mathbf{b}\rangle = \mathbf{I}_t^* \mathbf{h}\rangle$$

where the elements of the column matrix $\mathbf{h}\rangle$ are specified as

$$\dot{h}_n = \int_{-\beta}^{\beta} \dot{g}(\varkappa)\, \dot{F}_n^*(\varkappa)\, d\varkappa$$

Finally, on changing from the column matrix of expansion coefficients $\mathbf{b}\rangle$ to the excitation amplitudes of the array elements by

Eq. (13.22), we get

$$\mathbf{i}\rangle = \mathbf{\Pi}_t^* \mathbf{h}\rangle \tag{13.24}$$

In order to find the matrix product $\mathbf{\Pi}_t^*$, resort should be made to the condition of orthonormality for the system of functions $\langle \mathbf{e}(\varkappa)$ defined by Eq. (13.19), which may be written as a unit (identity) matrix of order N

$$\frac{1}{2\beta}\int_{-\beta}^{\beta} [\mathbf{e}(\varkappa)\rangle \langle \mathbf{e}^*(\varkappa)]\, d\varkappa = \mathbf{E} \tag{13.25}$$

In the above expression, the square matrix under the integral sign is obtained by multiplying a column by a row. The (np)th element of this matrix is the product $\dot{e}_n(\varkappa)\dot{e}_p^*(\varkappa)$ which, on integrating, yields either zero $(n \neq p)$ or unity. On substituting (13.20) into (13.25), we get

$$\frac{1}{2\beta}\mathbf{I}_t \int_{-\beta}^{\beta} [\mathbf{F}(\varkappa)\rangle \langle \mathbf{F}^*(\varkappa)]\, d\varkappa \mathbf{I}^* = \mathbf{I}_t \mathbf{R} \mathbf{I}^* = \mathbf{E}$$

where the elements of the $\mathbf{R}$ matrix are the integrals

$$r_{np} = \frac{1}{2\beta}\int_{-\beta}^{\beta} \exp[j\varkappa(z_n - z_p)]\, d\varkappa = \frac{\sin\beta(z_n - z_p)}{\beta(z_n - z_p)}$$

proportional to the mutual resistances of isotropic radiators.

On replacing $\mathbf{\Pi}_t^*$ by $\mathbf{R}^{-1}$ in Eq. (13.24), we obtain the final solution for the array synthesis problem based on the minimum mean-square-error approach

$$\mathbf{i}\rangle = \mathbf{R}^{-1}\mathbf{h}\rangle, \quad \dot{h}_n = \int_{-\beta}^{\beta} \dot{g}(\varkappa)\dot{F}_n^*(\varkappa)\, d\varkappa \tag{13.26}$$

There is no need to evaluate the orthonormal basis $\langle \mathbf{e}(\varkappa)$ in explicit form. A measure of incorrectness of the found solution may be given by the reactivity ratio

$$\gamma_n = \langle \mathbf{ii}^* \rangle / \langle \mathbf{iRi}^* \rangle$$

which is the ratio between the square of the current distribution norm and the active radiated power and which is similar to the reactivity ratio defined by Eq. (13.5).

Example 13.2. Let it be required to synthesize a radiation pattern such as would fit the delta-function as closely as possible

$$\dot{g}(\varkappa) = \delta(\varkappa - \varkappa_0)$$

where $\varkappa_0 \leqslant \beta$ is the direction of maximum radiation. On substituting this function into (13.26), we get

$$\dot{\mathbf{I}} \rangle = \mathbf{R}^{-1}\mathbf{F}^* (\varkappa) \rangle \qquad (13.27)$$

$$\mathbf{F}^* (\varkappa) \rangle = [\dot{F}_1^* (\varkappa), \dot{F}_2^* (\varkappa), \ldots, \dot{F}_N^* (\varkappa)]_t$$

Hence, the formal solution of the problem is

$$\dot{f}_N (\varkappa) = \langle \mathbf{F} (\varkappa) \dot{\mathbf{I}} \rangle = \langle \mathbf{F} (\varkappa) \cdot \mathbf{R}^{-1}\mathbf{F}^* (\varkappa_0) \rangle$$

The same result can be obtained if, in accordance with Eq. (11.31) we write the directivity of the array as

$$D = | \langle \dot{\mathbf{I}}\mathbf{F} (\varkappa_0) \rangle |^2 / \langle \dot{\mathbf{I}}\mathbf{R}\dot{\mathbf{I}}^* \rangle$$

and seek to obtain an excitation $\langle \dot{\mathbf{I}}$ such as would result in a maximum directivity. For this purpose, the expression for D is differentiated with respect to each $\dot{I}_n$ and the respective partial derivatives are equated to zero. The solution of the set of linear equations thus obtained will be exactly the same as given by (13.27). Thus, *the synthesis of a radiation pattern that would fit the delta-function as closely as possible is equivalent to the synthesis of a radiation pattern having a maximum directivity.*

By carrying out simple manipulations, it is possible to establish that the maximal directivity of an array made up of isotropic radiators in the direction $\varkappa_0$ is given by

$$D_{\max} = \langle \mathbf{F} (\varkappa_0) \mathbf{R}^{-1}\mathbf{F}^* (\varkappa_0) \rangle$$

If the array consists of directional elements, $\dot{F}_n (\varkappa)$ in the above equation may be replaced by $D_n^{1/2}\dot{F}_n (\varkappa)$, while also replacing r_{np} for isotropic elements by r_{np} for directional elements.

The choice of element spacing for an equally spaced array. In an array with a constant element spacing d, the positions of the elements are specified by the coordinates

$$z_n = (n - 1)\, d, \quad n = 1, 2, \ldots, N$$

and the array factor of the array

$$\dot{f}_N (\varkappa) = \sum_{n=1}^{N} \dot{I}_n \exp [j (n - 1) \varkappa d], \quad \varkappa = \beta \cos \theta$$

is a periodic function of the variable $\varkappa$

$$\dot{f}_N (\varkappa) = \dot{f}_N (\varkappa + 2\pi/d)$$

The ratio of the period $T = 2\pi/d$ to the extent of the visible range is $\lambda/2d$. If the radiators are spaced a distance $d = \lambda/2$ apart, the

period of the array factor is the same as the visible range. If the spacing is greater than a half-wavelength, more than one period of the function $\dot{f}_N(\varkappa)$ would occur within the visible range. Therefore, when one wishes to synthesize an arbitrary radiation pattern other than zero over the entire visible range, it is not justifiable to take the element spacing d greater than a half-wavelength because this would impair the accuracy of synthesis. On the other hand, if the element spacing d is less than a half-wavelength, the period of the function $\dot{f}_N(\varkappa)$ would be greater than the visible range, and the accuracy of synthesis would be enhanced owing to the use of a greater number of orthogonal functions $\dot{e}_n(\varkappa)$, provided the array length is held constant. However, the behaviour of the synthesized function $\dot{f}_N(\varkappa)$ beyond the visible range is uncontrollable, and, in consequence, the reactivity ratio might rise to prohibitively high values. Thus, so long as the element spacing remains smaller than a half-wavelength, a danger exists to obtain incorrect solutions with a high reactivity ratio. From a computational point of view, this lets itself felt in that the determinant of the **R** matrix tends to zero, and this leads to a loss of accuracy in finding the inverse matrix $\mathbf{R}^{-1}$. Therefore, care must be exercised in attempts to improve accuracy in the synthesis of a specified radiation pattern at the expense of a decreased element spacing.

13.7. Supergain Antennas

If we analyse the synthesis of an array having a maximum directivity on the basis of Eq. (13.27), with the number of elements N tending to infinity, with the element spacing d tending to zero, and with the array length held constant, $Nd = \text{const}$, we can see that $D_{\max}$ increases without bound. With any given antenna length it is possible to specify a number N of elements such that $D_{\max}$ will exceed any predetermined value. Alternatively, this result may be construed as a corollary of the statement given in Sec. 13.2 to the effect that in a linear antenna of any finite length it is possible to synthesize the array factor in the form of any arbitrary function. *This likely increase without bound in the directivity of an antenna having a finite length is called superdirectivity or, more commonly, supergain.* This occurrence stems from the fact that antenna synthesis from a specified pattern is an incorrectly stated problem.

A supergain antenna is likely to show the following.

(1) The current distribution in amplitude and phase becomes heavily 'ragged', and the phase of the excitation abruptly varies along the antenna (at a much faster rate than that of a speed-of-light wave).

(2) The amplitudes of the excitation at a fixed radiated power exceed substantially (by several orders of magnitude) the excitation amplitudes in an in-phase antenna of the same length and with the same radiated power.

(3) A huge amount of electromagnetic energy is stored in the near-zone field of a supergain antenna, and such an antenna is thus equivalent to a very perfect energy-storing element of negligible li- radiation losses. Matching of a supergain antenna to a transmission line becomes very problematic because, with an ideal reactive matching device, the antenna is equivalent to a resonant circuit of a colossal Q-factor (in excess of 10^5) and a negligible bandwidth. With a real matching device, its losses exceed radiation losses so much that the efficiency of a supergain antenna tends to zero.

One runs into similar consequences in an attempt to synthesize an arbitrary radiation pattern to an excessively high level of accuracy, for which reason it is legitimate to class as "supergain" all other incorrect solutions of the antenna synthesis problem.

Where may one draw a demarkation line between ordinary and supergain antennas? Arbitrarily, one may class as non-supergain all solutions of the synthesis problem, obtained by the Fourier integral method on the basis of Eqs. (13.7) through (13.10). As regards arrays, the Fourier integral method means that one should simply ignore mutual coupling between the radiators on setting $\mathbf{R} = \mathbf{E}$, where $\mathbf{E}$ is a unit matrix. Then, on the basis of Eq. (13.26), the amplitudes of the currents that energize the array elements are samples of the inverse Fourier transform

$$\dot{h}_n = \int_{-\beta}^{\beta} \dot{g}(\varkappa)\, \dot{F}_n^*(\varkappa)\, d\varkappa = \int_{-\beta}^{\beta} \dot{g}(\varkappa) \exp(-j\varkappa z_n)\, d\varkappa, \quad \mathbf{i}) \equiv \mathbf{h}) \qquad (13.28)$$

For the solutions of Eqs. (13.7) through (13.10) and (13.28) it is an easy matter to find the reactivity ratios which, as a rule, come out to be equal to unity very nearly. This means that for such nonsupergain antennas the excitation distribution is sufficiently smooth, the excitation amplitudes for a given radiated power is close to the least possible ones, and no appreciable amount of electromagnetic energy is stored in the near field. *Non-supergain antennas are most stable towards small errors in the excitation distribution*, which fact is essential from a practical point of view.

If a non-supergain antenna does not have a sufficient directivity or the specified radiation pattern cannot be synthesized with sufficient accuracy, but the dimensions of the antenna cannot be increased, it would be legitimate to realize to some extent the property of supergain by permitting a moderate increase in the reactivity ratio. For arrays, the entire class of solutions of the synthesis problem representing the minimum-mean-square-error match is defined

by the relation

$$|i\rangle = (\mathbf{R} + \alpha\mathbf{E})^{-1}|h\rangle$$

where $0 \leqslant \alpha \leqslant \infty$ is what we call the solution regularization parameter.

With α tending to infinity, the synthesis problem can be solved with the least attainable reactivity ratio. With α equal to zero, the other extreme case takes place — a solution without constraints on the reactivity ratio. However, numerous studies and calculations show that resort to the supergain condition does not hold any special promise from a practical point of view because one benefits too little in terms of directivity or the accuracy of radiation-pattern synthesis.

Chapter Fourteen

Aperture-Type Antennas

14.1. General Properties

The class of aperture-type antennas includes those in which one can, according to the equivalence principle, select a flat aperture surface S that can shape a pencil beam. The most common types in this class are the reflector (especially parabolic), horn, and lens antennas.

The directivity of aperture-type antennas is connected to the aperture surface S and the wavelength λ by a general equation

$$D = (4\pi S/\lambda^2)/K_{ae\Sigma} \qquad (14.1)$$

where $K_{ae\Sigma} \leqslant 1$ is the overall (resultant) aperture efficiency which depends on how the currents energizing the antenna are distributed in amplitude and phase, and on some other factors.

The — 3 dB width of the aperture antenna in any plane perpendicular to the aperture plane is inversely proportional to the linear dimension L of the aperture in that plane

$$\Delta\theta = (51°\lambda/L)\, K_{br\Sigma} \qquad (14.2)$$

where $K_{br\Sigma} \geqslant 1$ is the beam broadening ratio which depends on the aperture shape and the excitation distribution in amplitude and phase.

The distributors of aperture-type antennas are usually arranged so that the excitation power reaching each aperture element is delivered, in a geometrical-optical approximation, by an independent ray, with all the rays travelling the same distance from the common antenna input. Therefore, according to the feed system arrangement, aperture-type antennas may be classed into *shunt-fed*,

that is, those in which the elements of the radiating system are energized in parallel (in contrast to travelling-wave antennas where the elements are energized in series). An advantage of the shunt feed is that the aperture elements remain energized in phase and that the amplitude distribution is independent of the wavelength. Therefore, many aperture-type antennas, notably reflector antennas, are frequency-independent and can each operate in several frequency bands (from HF to EHF) at the same time. Of course, their directivity and beamwidth remain functions of wavelength, as follows from Eqs. (14.1) and (14.2).

It also follows from the above equations that the directivity of aperture-type antennas might increase without bound, and their beamwidth might tend to zero with increasing ratio S/λ^2. In real antennas, however, the maximum attainable directivity is limited because of random inaccuracies in antenna manufacture. With the present-day state of the art in antenna manufacture it is possible to build aperture-type antennas with a directivity of 10^9 and even more. Aperture-type antennas are superior to all other types in terms of directional properties.

14.2. Horn Antennas

Horn antennas, notably those in the form of a rectangular or circular waveguide open at the end, are widely used in the SHF and EHF bands. Radiators of this type may be used on their own or as feeds for lens and reflector antennas.

To begin with, let us consider a radiator in the form of an open end in a rectangular waveguide (Fig. 14.1). On reaching the open end, the dominant TE_{10} mode propagated in the guide is partly radiated and partly reflected back towards the generator. At the joint between the guide and free space, that is, at the aperture, higher modes are produced, and surface currents appear on the outer surfaces of the guide walls. For a semi-infinite circular waveguide, an exact solution to the electrodynamic problem of radiation into free space has been obtained by L.A. Wainstein of the Soviet Union. No exact solution has yet been found for the radiation field of a rectangular waveguide. In engineering calculations it is customary to deem that the tangential components of the field at the guide aperture (and, in consequence, the equivalent electric and magnetic

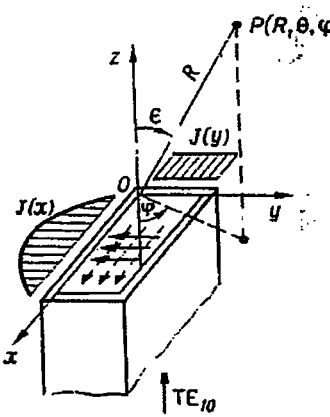


Fig. 14.1. Waveguide radiator

currents) are a sum of an incident and a reflected dominant wave

$$\dot{E}_y = -\dot{J}_x^m = (1 + \dot{\rho}) \sin(\pi x/a)$$
$$\dot{H}_x = -\dot{J}_y^m = \frac{[1-(\lambda/2a)^2]^{1/2}}{Z_m}(1 - \dot{\rho}) \sin(\pi x/a)$$

where $\dot{\rho}$ = reflection coefficient from the open end of the guide
$Z_m = (\mu_0/\varepsilon_0)^{1/2}$ = wave impedance of the medium

As has been found by experiments, in a rectangular waveguide with a side ratio of $b/a = 0.45$ the reflection coefficient is 0.2 to 0.3 and its phase is $\pi/2$ to $\pi/3$, the first figures applying to the upper frequency limit and the second to the lower frequency limit of the guide.

To a good degree of accuracy, the radiation pattern of the open end of the guide in the front hemisphere, as measured in the yOz plane (the E-plane) and in the xOz plane (the H-plane) is given by

$$F_E(\theta) = \frac{1+(\lambda/\lambda_g)\cos\theta}{1+\lambda/\lambda_g} \frac{\sin(0.5\beta b \sin\theta)}{0.5\beta b \sin\theta} \qquad (14.3)$$

$$F_H(\theta) = \left(\frac{\pi}{2}\right)^2 \cos\theta \frac{\cos(0.5\beta a \sin\theta)}{(\pi/2)^2 - (0.5\beta a \sin\theta)^2} \qquad (14.4)$$

Here, a and b are the inside dimensions of the waveguide, and λ/λ_g is the ratio of the free-space wavelength to the guide wavelength for the dominant TE_{10} mode. The angle θ is reckoned from the z-axis, and $\beta = 2\pi/\lambda$. A plot of the measured radiation patterns for a radiator in the form of the open end of a rectangular waveguide of dimensions 23 mm by 10 mm and operating at a wavelength of 3.2 cm is shown in Fig. 14.2.

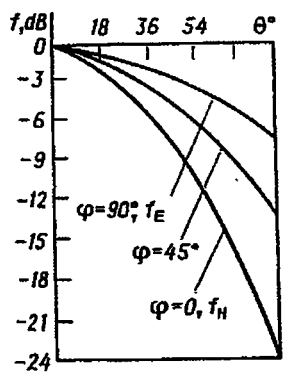


Fig. 14.2. Radiation pattern of the open end of a waveguide

Antennas in the form of the open end of a waveguide have low directional properties, and their gain usually ranges between 6 and 7 dB. Such antennas are most often used as phased array elements, as feeds for parabolic reflector antennas, and also as low-directional antennas on aircraft.

The directivity can be enhanced and reflections from the open end of the guide can be minimized by resort to horn radiators. Figure 14.3 shows an *H-plane sectoral horn*. This is a horn whose walls are flared in the H vector plane, and this is the reason for its name. The mode excited in the horn is similar to the TE_{10} mode in a rectangular waveguide. However, a sectoral horn differs from a guide in that its wave front forms a cylindrical surface, that the phase velocity is a variable quantity dependent on the ratio a/λ, and that

the field at a great distance from the horn throat takes the form of a purely transverse wave. The phase velocity is approximately given by

$$v = \frac{c}{[1-(\lambda/2a)^2]^{1/2}}$$

and comes very closely to that of light near the horn mouth, which fact serves to minimize reflections from the radiating surface of the aperture.

If the flare angle α_H is small, the wave front at the horn mouth is nearly plane, and one may use Eq. (14.4) to calculate the radiation

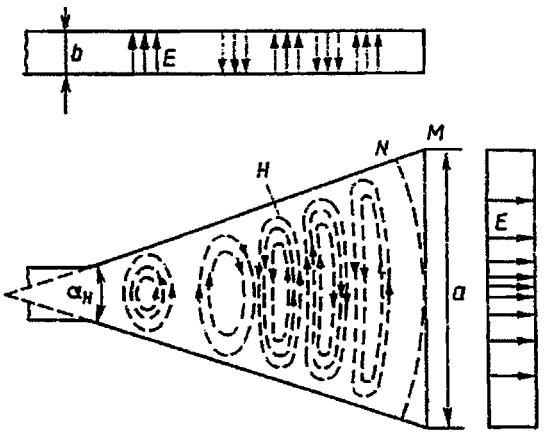


Fig. 14.3. H-plane sectoral horn and its field pattern

pattern in the *H*-plane. The main lobe of the radiation pattern is narrowed in the same ratio as the dimension a of the horn mouth increases in comparison with the broad dimension in a rectangular waveguide. As the flare angle α_H of the horn is increased, the wave front at the aperture becomes curved increasingly more, and this serves to broaden the radiation pattern. The phase of the field at the aperture edge as compared with its value at the aperture centre can be found by an approximate formula derived from geometrical construction (Fig. 14.3)

$$\Phi_2 = -(2\pi/\lambda)\, MN \approx -\pi a^2/4\lambda R = -(\pi a/2\lambda) \tan(\alpha_H/2) \quad (14.5)$$

where R is the length of the horn. The field phase at the horn aperture obeys a quadratic distribution. The effect of quadratic phase errors on the shape of the radiation pattern of a linear antenna has been examined in Chap. 11, and the radiation pattern plots shown in Fig. 11.9 hold for horn antennas as well.

As calculations show, the directivity of a horn antenna for a given horn length depends on the ratio a/λ in the way illustrated in Fig. 14.4. The presence of a maximum is explained by the fact that

an increase in the flare angle has two consequences. Firstly, it leads to an increase in the ratio a/λ, and this is responsible for a narrower radiation pattern. Secondly, as follows from Eq. (14.5), there is a fast increase in the phase error $|\Phi_2|$, and this is responsible for a broadening of the radiation pattern. Given a certain electrical dimension of the horn aperture, the joint action of the two factors produces a maximum directivity. It turns out that, *given any horn length, the directivity will have a maximum when the quadratic phase*

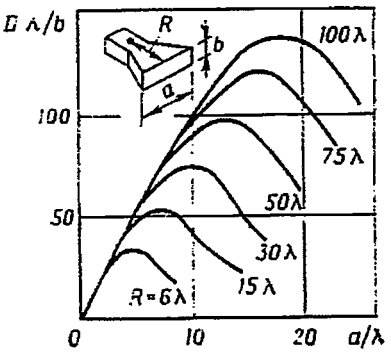


Fig. 14.4. Directivity of an H-plane sectoral horn as a function of dimensions

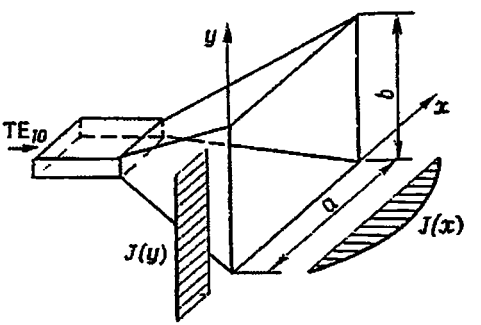


Fig. 14.5. Pyramidal horn

error at the edge of the horn is 135°. The H-plane sectoral horn satisfying this condition is referred to as the optimum one. The total aperture efficiency of an optimum H-plane sectoral horn is around 0.64 (the figure is 0.81 because the amplitude distribution tapers to zero at the horn edges, and 0.79 due to a quadratic phase error).

Besides H-plane sectoral horns, resort is made to E-plane sectoral horns whose walls are flared in the E vector plane. The width of the radiation pattern in the H-plane of an E-plane sectoral horn is the same as it is at the open end of a waveguide, whereas in the E-plane the beamwidth decreases with increasing b, if the flare angle α_E is taken small enough. In an E-plane sectoral horn the field amplitude distribution at the aperture is nearly uniform and the quadratic phase error at the aperture edge, $|\Phi_2| \approx -\pi b^2/4\lambda R$, corresponding to an optimum horn with a maximum directivity is 90°. When $|\Phi_2|$ is less than 45°, the radiation pattern in the E-plane can be evaluated by Eq. (14.3). The reduction in the aperture efficiency in the E-plane because of the quadratic phase error can be found from the plot in Fig. 11.10.

Most often, use is made of *pyramidal horns* of rectangular cross-section. A pyramidal horn is one that is flared in both the E and H planes (Fig. 14.5). This type of horn has a sharp beam in both planes

but the beam in one plane can be adjusted independently of the other. A spherical wave is excited in a pyramidal horn, with a variable phase velocity which comes very closely to that of light near the open end of the horn. Because of this, reflections from the aperture are negligible — the horn fully matches the waveguide to free space. The field phase errors in the aperture can be found by Eq. (14.5) in the H-plane and, by a similar equation (on replacing a by b), in the E-plane. In the case of small phase errors (with $|\Phi_2|$ less than 45°), the radiation pattern of a pyramidal horn differs only slightly from that of an in-phase rectangular aperture having a suitable amplitude distribution and can therefore be evaluated by Eqs. (14.3) and (14.4). As a way of reducing the horn length, it is legitimate to allow a quadratic phase error in the field phase at the aperture to be as great as $|\Phi_2| = 135°$ in the H-plane and $|\Phi_2| = 90°$ in the E-plane. As has been noted, such a horn would be optimal and its aperture efficiency would roughly be given by

$$K_{ae} = K_{ae.a}K_{ae.p}^2 = 0.81 \times 0.8^2 = 0.52$$

where the first factor takes care of variations in the amplitude distribution in the H-plane, and the second, of quadratic phase errors in the E and H planes.

There are also horns which are of conical rather than rectangular cross-section. They are known as *conical horns*. They are formed by flaring the open end of a circular guide excited by a TE_{11} wave. The radiation of a conical horn is similar to that of a pyramidal horn, and its optimal dimensions may be regarded as standing mid-way-between those of optimal E- and H-plane sectoral horns.

Among the advantages of horn antennas are simple construction and a satisfactory performance over the bandwidth. Practically all optimal and longer horns may be used over the entire bandwidth of the feed waveguide. When used on their own, horn antennas are most often employed for measurement purposes, such as standard antennas having a known gain. Also, horns are widely used as feeds for reflector and lens antennas and in some other types such as impedance antennas.

14.3. Lens Antennas

Lens antennas depend for their operation on the optical properties of electromagnetic waves because the dimensions and radius of curvature of a lens antenna are usually many times the wavelength. It is typical of lens-type antennas that they modify a cylindrical or a spherical wave front into a plane wave front. Figure 14.6 shows a dielectric lens antenna in which a feed F launches a spherical wave. Refracted by the convex surface of the lens, the spherical wave front

is transformed into a plane one; at the same time the wavelength in the dielectric is reduced as

$$\lambda_d = \lambda/\sqrt{\varepsilon}$$

The exit surface of the lens is flat, and the wave front on leaving it remains a plane one. Because the dimensions of the exit surface are great in comparison with the wavelength, the resultant radiation is a sharp beam.

Define the contour of the lens. Let the ray emerging from point F arrive at point P on the surface of the lens as shown in

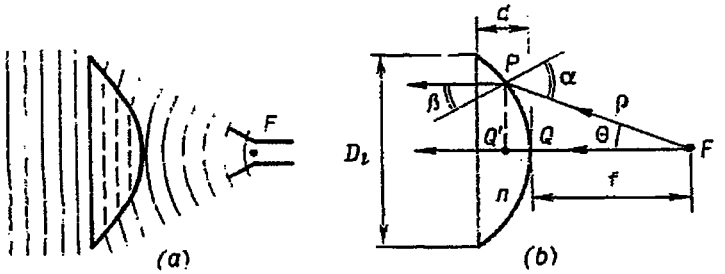


Fig. 14.6. Dielectric lens: *a*, general arrangement; *b*, ray paths in it

Fig. 14.6*b*, such that the angle of incidence is α and the angle of refraction is β. These angles are reckoned relative to a local normal to the illuminated surface of the lens. By *Snell's law of refraction of rays*

$$\sin \alpha = n \sin \beta$$

where

$$n = \sqrt{\varepsilon}$$

is the *index of refraction* (or *refractive index*) of the dielectric. The equation describing the contour of a lens can be derived on the basis of what is known as the *equal-path-length law* which states that since wave fronts are equiphase surfaces, it follows that between any two specific wave fronts (original and transformed) the optical (electrical) path lengths along all ray paths must be equal, that is,

$$FP = FQ + nQQ'$$

Hence,

$$\rho = f + n(\rho \cos \theta - f)$$

or

$$\rho = (n - 1) f/(n \cos \theta - 1) \qquad (14.6)$$

where f = focal distance, or the distance from the focal point F to the vertex of the lens, Q

θ = angle between the lens axis and the ray
ρ = distance from the focal point to the current point on the surface of the lens

Equation (14.6) is the equation of a hyperbola. Thus, the *convex surface of a dielectric lens must be a hyperbola in contour*. The lens thickness d is determined over the optical path and is connected to the lens diameter D_l by a relation of the form

$$d = (1/n)\,(D_l/2 \sin \theta_{max} - f) \qquad (14.7)$$

where θ_{max} is the angle between the axis of the lens and the direction from the focal point to the extreme point of the lens contour. The angle $2\theta_{max}$ at which the edges of the lens are seen looking from the focal point is often called the *aperture angle* of a lens. Equations (14.6) and (14.7) enable one to calculate the contour and geometrical dimensions of any lens.

Because the exit surface of a lens is an in-phase surface, it may be regarded as a flat (planar) radiating aperture for which the lens itself and its feed make up the feeder system (a distributor). With a quasi-point source feed (which may be a half-wave dipole), the lens may be a body of revolution, and it will then have a circular aperture. With a line source feed located on the focal line F, the lens must be cylindrical, and its aperture may be rectangular in shape. Dielectric lenses are broadband devices if their refractive index is frequency-independent. In the SHF band such lenses are used but seldom because they turn out to be too heavy in weight and difficult to make. Interest in dielectric lenses has increased since the advent of operation at extra-high and even higher frequencies (millimetric and shorter wavelengths) where the absolute dimensions of a lens are small.

Attempts to reduce the weight of lenses in which the electrical (optical) path length is increased by the lens medium (these are so-called *delay lenses*) have led to the development of artificial dielectrics — an array of discrete metal particles (spheres or discs substantially smaller than a wavelength in diameter) or metal ribbons supported in a matrix of a lightweight foamed dielectric. Because of the difficulties in their manufacture and serious limitations in their design, artificial-dielectric lens antennas are used on a limited scale.

A much wider use is made of what are known as *metal-plate lens antennas*, in which the phase velocity is increased. This is the reason why they are called *phase-advance lenses*. This type of lens is made up of a stack of parallel metal plates which are spaced some distance a apart and form a concave surface (Fig. 14.7). The vector of the electric field E excited by a feed located at the focal point of the lens should be parallel to the plates. Then the space between two adjacent plates will act as a *waveguide* propagating a TE_{10} wave with a phase

velocity given by

$$v = \frac{c}{[1-(\lambda/2a)^2]^{1/2}}$$

Thus, a medium exists between the plates for which the effective refractive index is

$$n = c/v = [1 - (\lambda/2a)^2]^{1/2} < 1 \qquad (14.8)$$

For the wave propagated between the plates to be solely the TE_{10} mode, the plate separation a must be such that $\lambda/2 < a < \lambda$. In consequence, the refractive index for a phase-advance lens lies in the range $0 < h < 0.75^{1/2}$.

The equation of the surface contour for a metal-plate lens can be derived, using the construction shown in Fig. 14.7*b*. The ray emer-

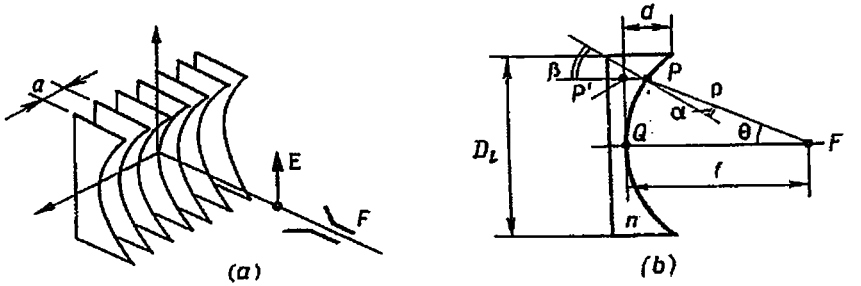


Fig. 14.7. Metal-plate lens: *a*, general arrangement; *b*, ray paths in it

ging from the focal point F and arriving at point P on the illuminated surface of the lens is refracted as given by Snell's law of refraction of rays

$$\sin \alpha = n \sin \beta$$

The optical path FQ must be equal to the optical path FP', that is,

$$FQ = FP + nPP$$

This condition again leads to Eq. (14.6) defining the contour of a lens, but at $n < 1$ this is the equation of an *ellipse*. The thickness of the lens contour is given by

$$d = (1/n)\,(f - D_l/2 \sin \theta_{max})$$

where $2\theta_{max}$ is the aperture angle of the lens and D_l is the aperture diameter.

A phase-advance lens assembled from elliptically contoured identical plates will focus the radiation in the E vector plane. If, on the other hand, a phase-advance lens is made up of rectangular plates differing in the value of d, so that they, too, form an elliptical contour, the lens will focus the radiation in the H vector plane. In

such a case, the angle of refraction of the ray at the illuminated surface of the lens will be independent of the refractive index n, and the direction of the refracted rays will, at any value of n, be determined by the direction of the waveguides between the metal plates. Quite aptly, such lenses are called *constrained lenses*. For the radiation to be focused in both the E and the H planes, a metal-plate lens must be made up of suitably contoured plates differing in thickness.

Small and medium-sized metal-plate lenses are very simple in design. When used in conjunction with a horn, the length of a metal-plate lens can be cut down appreciably. For example, if we wished to

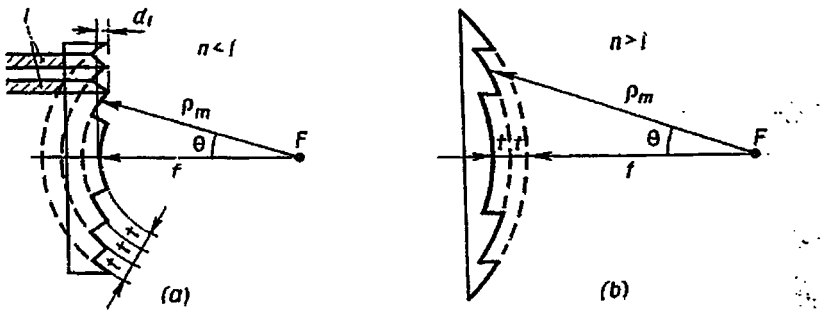


Fig. 14.8. Zoned lenses: *a*, phase-advance type; *b*, delay type

obtain an optimum horn with an aperture dimension $a = 20\lambda$, its length would have to be taken equal to $R = 200\,\lambda$. With a lens placed at its mouth, the horn may have a length of the same order of magnitude as the aperture dimension, that is, $R = 20\,\lambda$. A serious disadvantage of metal-plate lenses is their limited bandwidth because, as follows from Eq. (14.8), the refractive index depends on the wavelength. Calculations show that if the refractive index at the centre frequency be taken equal to 0.5 and if we allow a quadratic phase error equal to $|\Phi_2| = \pi/2$ at the extreme frequencies, the relative bandwidth (in per cent) will be

$$2\Delta f/f_0 = 33\lambda_0/d$$

At d substantially greater than the free-space wavelength λ_0, the bandwidth comes out to be too limited. The bandwidth of a metal-plate lens antenna can be increased and its thickness d decreased by resort to what is known as *zoning*. The principle of zoning is illustrated in Fig. 14.8. The illuminated surface of the lens facing the feed has a stepped cross-section, with the thickness of a zone step t being such that the rays refracted by adjacent zones of the lens could arrive at the lens aperture with a phase difference of 2π. Each zone has a focal length of its own, and the equation for the surface

contour of the mth zone takes the form

$$\rho_m = (1 - n) f_m/(1 - n \cos \theta)$$

where

$$f_m = f + (m - 1) t, \quad m = 1, 2, 3, \ldots$$

The dimension t is chosen such that at the design wavelength the field at the aperture is in-phase

$$\beta_0 t - \beta_0 n t = 2\pi$$

Hence,

$$t = \lambda_0/(1 - n)$$

Zoning is responsible for the occurrence of unilluminated or shadow areas in the lens aperture (Fig. 14.8). The shadow areas reduce the aperture efficiency because some of the radiated power is scattered by the zone steps. Given the allowable quadratic phase error $|\Phi_2| = \pi/2$ and a refractive index equal to $n_0 = 0.5$, the bandwidth (in per cent) of a zoned phase-advance lens will be

$$2\Delta f/f_0 = \frac{50}{M - 1 + 1.5\,(d_1/\lambda_0)}$$

where the contour thickness d_1 is shown in Fig. 14.8*a*, and M is the number of zones in the lens. The quantity d_1 is connected to the thickness d of an unzoned lens (see Fig. 14.7*b*) as

$$d_1 = d - (M - 1) t$$

and for $t = 2\lambda_0$ (for $n_0 = 0.5$) the equation takes the form

$$2\Delta f/f_0 = \frac{50}{1.5\,(d/\lambda_0) - 2\,(M - 1)}$$

Thus, resort to zoning with a fixed value of d substantially extends the bandwidth of a phase-advance (metal-plate) lens.

Zoning might also be used to reduce the thickness and mass of a delay lens (see Fig. 14.8*b*), but this would be accompanied by a reduction in per cent bandwidth due to the occurrence of quadratic phase errors in accordance with the equation

$$2\Delta f/f_0 = 50/(M - 1)$$

and a reduction in the aperture efficiency due to the appearance of shadow areas in the aperture.

Let us take a closer look at the amplitude distribution in the aperture of a lens antenna. As an example, consider a delay lens and a phase-advance lens isotropically illuminated from the focal point (Fig. 14.9*a* and *b*). The section-lined bundles of rays in the figures account for an equal fraction of the total power flow because the bundles subtend the same angle or, which is the same, have the

same angular width

$$\Delta\theta_1 = \Delta\theta_2$$

In a delay lens, the outer bundle, after it has been refracted, broadens, and this corresponds to a reduction in the radiation intensity at the edge of the lens. The picture is entirely different in a

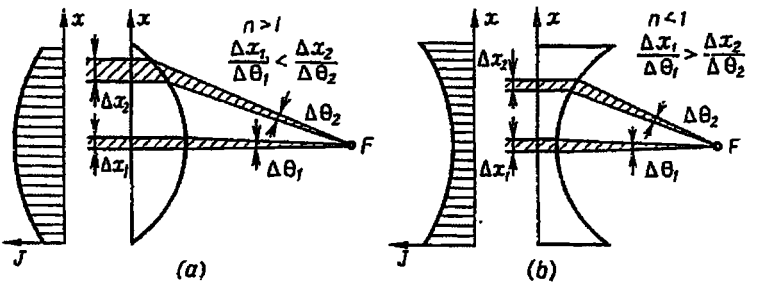


Fig. 14.9. Amplitude distribution in lenses

phase-advance lens. Here, the outer bundles, after they have been refracted, are compressed, and the amplitude distribution tends to rise towards the lens edge.

The above factors are not decisive in the shaping of the amplitude distribution. What counts most of all is the radiation pattern of the feed — its shape governs the distribution of power among the individual ray bundles. This explains why it is to meet especially stringent requirements. Within the aperture angle of a lens the radiation pattern of the feed should make up for the errors arising in the amplitude distribution due to the lens. At $|\theta| > |\theta_{max}|$, the radiation from the feed should be as small as possible so as not to produce a power flow that would miss the lens and contribute to the sidelobe radiation of the antenna — an event known as *spillover*, when some of the feed power spills over the aperture edges. It is important for the lens feed to have a phase centre coincident with the focal point of the lens since otherwise phase errors might arise in the aperture. An idea about an optimal shape of the feed radiation pattern can be gleaned from Fig. 14.10 where the section-lining shows the portion of the radiation pattern that does not contribute to the aperture illumination and is responsible for spillover.

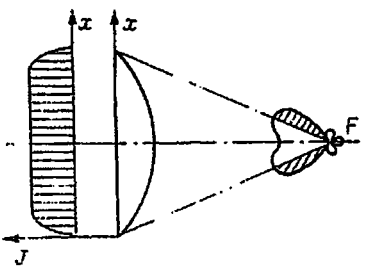


Fig. 14.10. Equalization of the amplitude distribution through the use of a feed with a suitably shaped radiation pattern

A further amount of power is scattered due to unwanted reflections from the refracting surfaces of the lens. These reflections can be minimized through the use of 'transparent' (matching) quarter-wavelength layers for which the refractive index is equal to the square root of the refractive index of the lens medium. Alternatively, this can be done by other methods that can provide reflection compensation. In lens antennas with an optimum feed the overall aperture efficiency may be as large as 0.80 to 0.85. However, large lens antennas are too unwieldy as compared with reflector antennas — a fact which explains why the simpler reflector antennas are preferable in most applications, especially when the aperture has great electrical dimensions (in excess of 50λ).

Luneberg lens antennas. Unique among all lens antennas are those made up of a nonuniform dielectric possessing spherical

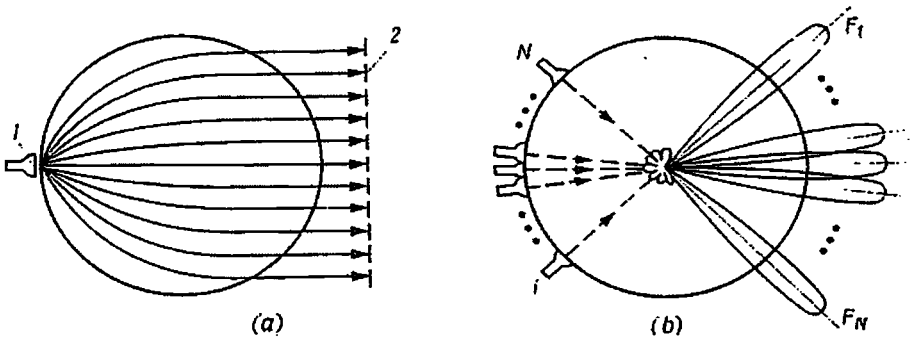


Fig. 14.11. Spherical Luneberg lens

symmetry. They are called *Luneberg lens antennas* after their inventor who showed that if a dielectric sphere has a refractive index which varies with the distance from its centre as

$$n = [2 - (R/R_{max})^2]^{1/2} \qquad (14.9)$$

where R_{max} is the radius of the sphere and R is the current radius of a point inside the sphere, such a lens will modify the spherical wave front from a point source feed, *1*, located at the surface of the sphere into a plane wave front, *2*, diametrically opposite to the feed (Fig. 14.11*a*). By moving the primary source feed (such as the open end of a waveguide or a horn) on the lens surface, it is possible to scan the beam in all directions without any deterioration in the radiation pattern. Most often, however, spherical Luneberg lenses are fitted with a cluster of feeds each of which forms a pencil-beam radiation pattern of its own (Fig. 14.11*b*). Such a structure is known as a *multibeam antenna system*. It can serve, independently and at the same time, several transmitters or receivers. A Luneberg lens anten-

na also provides a means for continuous radar surveillance within a wide sector of azimuth angles and also to carry on point-to-point radio communication with many distant stations.

Physically, a Luneberg lens may be an assemblage of uniform concentric layers of foamed r.f. ceramics, with the density of the layers decreasing from centre to periphery so that the refractive index could vary as given by Eq. (14.9).

In addition to the spherical type, there are cylindrical Luneberg lenses which permit the beam to be scanned without pattern deterioration in only one plane.

14.4. Parabolic Reflector Antennas

Reflector antennas are the most commonly used ones of all pencil-beam antennas. Their wide use in a great variety of radio systems is due to simple design, the possibility of obtaining various radiation patterns, high efficiency, a low noise temperature, a satisfactory bandwidth, and some other advantages. In radar applications, reflector antennas offer means for readily obtaining an equisignal zone

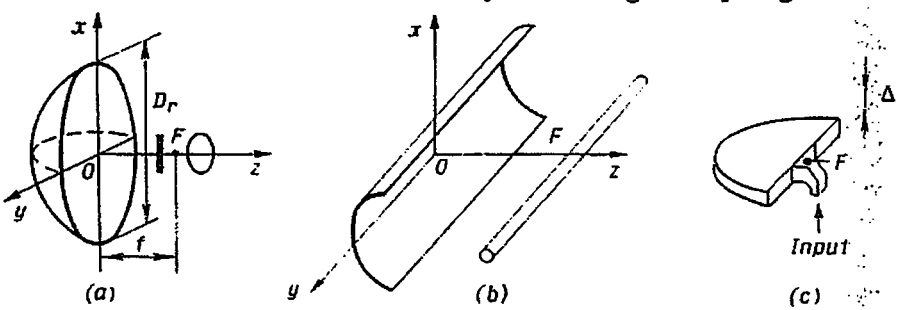


Fig. 14.12. Parabolic reflector antennas: *a*, paraboloid of revolution ('dish'); *b*, parabolic cylinder; *c*, pillbox or cheese

and producing several radiation patterns by a common reflector (including sum and difference patterns). Some types of reflector antennas allow the beam to be slewed at a high rate over an appreciable sector in azimuth and/or elevation. Reflector antennas are the most commonly used type in space communications and radio astronomy — it is with the aid of reflector antennas that there is a possibility to build mammoth antenna systems with an effective aperture area of many thousands of square metres.

The classical types are *parabolic reflector antennas* which can be built in the form of a paraboloid of revolution, a parabolic cylinder, or a structure bounded by parallel conducting surfaces, popularly called *cheese* or *pillbox antennas* (Fig. 14.12). A paraboloidal reflector, known more commonly as a *dish*, is illuminated by a feed of low

directional properties (say, a horn) placed at the focal point of the dish; it converts a spherical wave front into a plane wave. A *parabolic-cylinder* (or *cylindrical parabolic*) reflector is illuminated by a linear antenna placed on a line passing through the focus (the focal line); it converts a cylindrical wave front into a plane wave. As is the case with lenses, these antennas use the optical properties of radio waves. The geometric properties of a parabola are such that the rays emerging from the focal point and reflected from the parabolic surface are collimated into a beam of rays parallel to its axis. The path length from the focus to the parabola and then to the aperture line passing through the edges of the parabola is the same for any angle θ (Fig. 14.13). Thus, an in-phase surface is formed in the aperture of a parabolic antenna, and a pencil beam of radiation is produced.

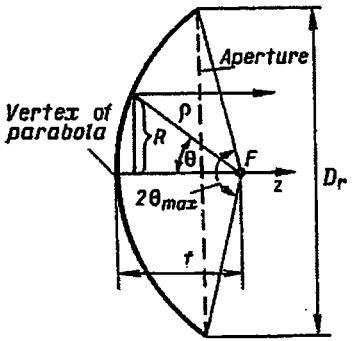


Fig. 14.13. Contour of a parabolic reflector antenna

With the origin coordinates placed at the vertex, the equation of a paraboloid of revolution is

$$x^2 + y^2 = 4fz$$

whereas in spherical coordinates (with the origin of coordinates placed at the focus) the equation of a paraboloid of revolution is

$$\rho(\theta) = 2f/(1 + \cos\theta)$$

There are long-focal-length and short-focal-length parabolic reflector antennas. In a *long-focal-length device*, the focal point is located outside the antenna, $f > D_r/4$, and the aperture angle $2\theta_{max}$ at which the reflector edges are seen looking from the focal point is $2\theta_{max} < \pi$. In a *short-focal-length antenna*, the focal point lies between the reflector and its aperture, such that $f < D_r/4$ and $2\theta_{max} > \pi$.

The radiation of parabolic reflector antennas is traditionally calculated by the method of equivalent surface electric and magnetic currents. The radiation integral can be written in any one of two forms:

(1) With integration carried out over the surface area of the flat aperture and over the shadow (rear) side of the paraboloid.

(2) With integration carried out over the curvilinear surface which extends in on the illuminated and the shadow sides of the paraboloid.

For simplicity, the relatively small electric surface currents on the shadow side of the paraboloid are neglected in each case.

In the first case, known as the *aperture field method*, the aperture is flat, and the equivalent electric and magnetic currents are deemed to be in phase with one another, and their amplitude distribution is

assumed to vary as the ratio $F(\theta, \varphi)/\rho(\theta)$, where $F(\theta, \varphi)$ is the field-strength radiation pattern of the feed and $\rho(\theta)$ is the distance from the focus to a point on the surface of the paraboloid. With this approach, the theory of a flat (planar) aperture (see Sec. 12.3) is fully applicable.

In the second case, known as the *induced current method*, the illuminated side of the surface is exactly coincident with the surface of the paraboloid, and so the equivalent magnetic surface currents are zero whereas the surface electric currents are physical currents flowing on the illuminated side of the paraboloid. The electric surface current is approximately defined as twice the tangential component of the magnetic field intensity due to the wave emerging from the feed. The amplitude and phase distribution of the induced currents in this case is a far more complex function of coordinates than it is in the former case; also, the direction of currents changes from one point on the surface of the paraboloid to another.

Neither of the two above techniques is exact because the surface current distribution is determined approximately. Still, the aperture field method is less accurate because the geometrical-optics approximation under which the amplitude-phase distribution in the aperture is determined is valid only with the wavelength tending to zero. In consequence, this method may be used if the reflector diameter D_r is substantially greater than the wavelength λ. The induced current method is more accurate but it is more elaborate. The recent trend has been towards more rigorous approaches in the calculation of radiation from reflector antennas, notably those based on the *geometrical theory of diffraction*, which allow the designer to determine to a satisfactory degree of accuracy the radiation pattern not only in the region of the main lobe, but also in the region of sidelobe and backward radiation. These approaches are used in the now widely spreading computer-aided design of antenna systems.

Consider the distribution of the currents induced by a feed on the illuminated side of a paraboloid of revolution. Let the feed be an electric dipole oriented along the x-axis and fitted with a flat subreflector in the form of a disc (Fig. 14.12*a*). The magnetic field vector lies in the yOz plane, so on the surface of the paraboloidal reflector the magnetic intensity is

$$\dot{H}(\theta, \varphi) \equiv \dot{F}(\theta, \varphi) \exp(-j\beta\rho)/\rho(\theta)$$

where $\dot{F}(\theta, \varphi)$ is the radiation pattern of the feed in spherical coordinates with centre at the focal point of the paraboloidal reflector. The surface electric current density vector is

$$\mathbf{J}^e \equiv [\mathbf{H}, \mathbf{n}]$$

and since the unit vector normal to the surface of the paraboloidal reflector has all the three components, then the surface electric cur-

rent density vector should likewise have three components, namely $\dot{J}^e_x$, $\dot{J}^e_y$ and $\dot{J}^e_z$. The applicable electric current distribution pattern projected onto the xOy plane is shown in Fig. 14.14*a*. In all the four quadrants the current components retain a general vertical orientation, but the relatively small horizontal components change direction from one quadrant to another.

For a feed in the form of a magnetic dipole with a vertically polarized **E** vector (this may be a half-wave slot cut in a small sheet parallel to the y-axis), the current distribution on the reflector will pre-

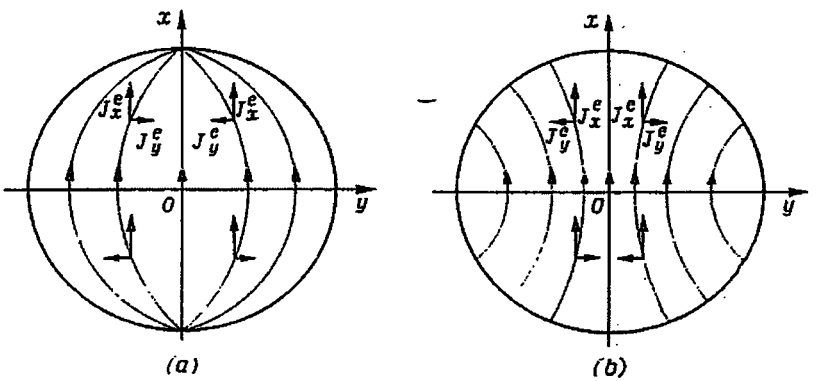


Fig. 14.14. Surface current distribution in the aperture of a paraboloidal reflector

sent a different structure (Fig. 14.14*b*). But here, too, the horizontal current components change direction from one quadrant to another.

In the two principal planes, xOz and yOz, the radiation pattern of a parabolic antenna is decided by the radiation due to the vertical current component $\dot{J}^e_x$, whereas the radiation due to the $\dot{J}^e_y$ component in these two planes will cancel out. The field radiated in the two principal planes is linearly and vertically polarized. In other planes passing through the z-axis, an additional contribution comes from the radiation due to the $\dot{J}^e_y$ components, and this is responsible for the occurrence of a parasitic horizontal polarization of the radiated field. The overall field turns out to be elliptically polarized. The parasitic horizontal cross-polarization brings on a reduction in the total aperture efficiency of the antenna. The cross-polarization level varies inversely as f/D_r, or inversely as the focal length of the paraboloid.

The choice of parameters for a parabolic reflector antenna. When a parabolic reflector antenna is calculated by the aperture field method, it is customary to assume that the radiation pattern of the feed, $F(\theta)$, possesses rotational symmetry about the axis of the

reflector. Using the known radiation pattern of the feed, the amplitude distribution in the aperture can then be found by the following equations:

$$I(\theta) = F(\theta)/\rho(\theta) = F(\theta)(1 + \cos\theta)/2f$$
$$\sin\theta = R/\rho(\theta) = R(1 + \cos\theta)/2f$$

where R = radial coordinate in the aperture plane
θ = angle between the axis of the paraboloid and the direction from the focus to the current point of reflection on the paraboloid (see Fig. 14.13)

The amplitude distribution thus found is approximated by Eq. (12.12), which means that one finds a suitable value of the power index n and the quantity $(1 - \Delta)$ which gives the field level at the aperture edge. Carrying out the necessary calculations by Eqs. (12.13) and analysing the results, one then can determine the beamwidth of the antenna and the sidelobe level. The aperture efficiency K_{ae} which takes care of variations in the amplitude distribution is then found by Eq. (12.8) or by looking up an aperture distribution table in a reference handbook.

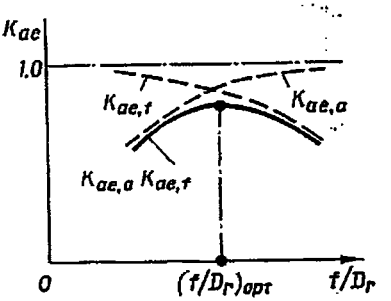


Fig. 14.15. Area efficiency as a function of focal length

On carrying out simple integration, one can also determine the feed efficiency K_f, that is, the fraction of its radiated power reaching the reflector

$$K_f = \int_0^{\theta_{max}} F^2(\theta)\sin\theta\,d\theta \Big/ \int_0^{\pi} F^2(\theta)\sin\theta\,d\theta < 1 \qquad (14.10)$$

This quantity accounts for the spillover, or the loss of the feed power as it 'spills' over the reflector edge.

The parameters of the reflector antenna system thus found come out to be dependent on the relative focal length f/D_r or, which is the same, the aperture angle of the reflector $2\theta_{max}$. A typical plot of K_{ae} and K_f as functions of f/D_r is shown in Fig. 14.15 by the dashed curve, while the full curve represents their product, that is, the overall aperture efficiency. As is seen, with a given feed, there is an optimal value of the relative focal length, $(f/D_r)_{opt}$, at which the overall aperture efficiency is a maximum. At low values of the relative focal length, the feed efficiency tends to unity, but the radiation pattern of the feed is narrow as compared with the aperture angle of the reflector, and the overall aperture efficiency decreases due to a very nonuniform amplitude distribution (the illumi-

nation strongly tapers down towards the edges). If the relative focal length is great, the amplitude distribution is made more uniform and the aperture efficiency increases, but the reflector intercepts only a part of the total power supplied by the feed in order to shape the required radiation pattern whereas the rest is wasted and only serves to increase the scatter fraction of the antenna.

Calculations have shown that when the feed is in the form of a half-wave dipole fitted with a subreflector, the optimal relative focal length is $(f/D_r)_{opt} = 0.38$, and the illumination level at the aperture edge in the optimal case is around 0.33 (or —10 dB) relative to the centre of the aperture. In the circumstances, the overall aperture efficiency is 0.83. If the feed produces a sharper radiation pattern than does a half-wave dipole with a subreflector, the optimal relative focal length increases. Conversely, if the feed produces a broader radiation pattern, the reflector should have a shorter focal length. Roughly, it may be taken that the optimal illumination level at the reflector edge should be —10 dB relative to the centre of the aperture. The beamwidth (in degrees) is given by

$$\Delta\theta = (65 \text{ to } 70)\,\lambda/D_r$$

and the level of the first (greatest) side lobe is approximately 22 to 24 dB down relative to the main lobe.

The approximate method discussed above is used in the sketch design of inexpensive reflector antennas of small size (with a reflector diameter of around 2 or 3 m for the SHF band). In this case, a small loss in the aperture efficiency because the design is other than optimal can be made up for by an increase in the aperture size. Unfortunately, the cost of large reflectors usually increases as the second power of the diameter or even faster. Therefore, in the case of a large reflector antenna it is essential to obtain a maximal aperture efficiency, and the entire reflector antenna and its feed must be optimized through more exact calculations which would take into account the greatest possible number of relevant factors. As a rule, the radiation pattern of the feed should be synthesized by a special procedure so as to obtain the largest value for the overall aperture efficiency of the antenna.

The most important factors affecting the overall aperture efficiency of a reflector antenna and its noise temperature (as mediated by the power scatter into the side lobes of the radiation pattern) may be summed up as follows.

1. The form of the amplitude distribution in the aperture and the corresponding aperture efficiency $K_{ae,a}$.
2. The feed efficiency K_f defined by Eq. (14.10).
3. The blockage of some of the aperture area by the feed and its supporting structure. Aperture blocking is accounted for by intro-

ducing the blockage factor

$$K_b \approx (S - S_b)/S$$

where S is the overall surface area of the aperture and S_b is the area blocked by the feed and its support structure. Feed blockage leads to a rise in the scatter fraction and, in consequence, to a higher noise temperature of the antenna.

4. Spillover of electric currents onto the shadow side of the reflector, thus boosting the far-out side lobes. This effect is accounted for by the diffraction factor K_{dif} which is found by techniques from diffraction theory.

5. Cross-polarization accounted for by a suitable factor, K_{cp}.

6. Phase errors (systematic and random) in the distribution of equivalent electric currents in the aperture. Systematic errors (such as the non-spherical shape of the phase characteristics of the feed, inaccuracies in the placement of the phase centre at the focal point of the reflector) may be eliminated by improvements in its engineering design and by alignment. Random errors (deviations from the required reflector contour, strains induced in the structure by its weight and winds, variations in dimensions with temperature, and the like) are accounted for by a suitable factor, K_σ. Random errors limit the maximum size of an antenna and the upper limit of the frequency band over which the reflector can be operated (see later).

7. Ohmic power losses due to the finite conductivity of the reflecting surfaces, the presence of non-ideal dielectrics in the feed, anti-corrosion coatings, and so on. These losses affect the antenna efficiency η_A and bring on a rise in the noise temperature by the equation

$$T_n = T_0 (1 - \eta_A)$$

where $T_0 = 273$ K.

8. The effect of the radio-transparent enclosure (if one is used), as it gives rise to additional phase errors in the current distribution in an equivalent aperture, and also causes additional power attenuation as radio waves pass through the enclosure ([11], vol. 1).

To sum up, the performance of a reflector antenna depends on a great number of factors interrelated in an intricate manner. For example, an attempt to make the amplitude distribution more uniform and to enhance the feed efficiency would require an increase in the electrical dimensions of the feed, but this would lead to an untoward increase in antenna size. As a rule, the overall area efficiency of a parabolic reflector antenna, which is the product of all partial efficiencies in turn affected by all of the factors listed above, is a mere 0.5 to 0.65, rising to 0.75 in the best of cases. Where a large reflector antenna is involved (with a reflector diameter of more than 10 m), another important consideration besides the area effi-

ciency is the overall cost of the structure. That is why a particular design for a large reflector antenna is chosen by considering in minute detail the results of many preliminary calculations.

Tolerances on the reflector of a parabolic antenna. Let us denote as Δ the random deviation of the reflector contour from a parabola. Assume also that this quantity is normally distributed with a mean of zero and a variance $\overline{\Delta}^2$. Then the maximum deviation of the reflector contour from the design value will not exceed

$$\Delta_{max} = 2.6\overline{\Delta}$$

with a probability of 99%. Customarily, the quantity Δ_{max} is referred to as the tolerance on the surface contour of a parabolic reflector.

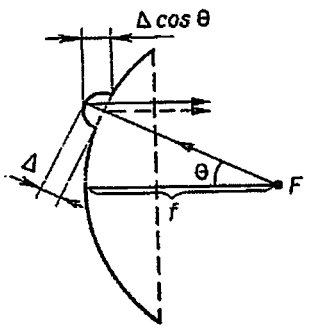


Fig. 14.16. To calculating the phase error at the aperture of a reflector

On recalling that the path length travelled by the radiated power to the aperture (more accurately, from the feed to the reflector and then from the reflector to the aperture, Fig. 14.16) varies in a random manner, we may write an approximate expression for the rms value of a random phase error in the aperture as

$$\overline{\Phi} \approx 2\beta\overline{\Delta} = 4\pi\overline{\Delta}/\lambda$$

The correlation radius for discontinuities on the surface of the reflector is usually several times the wavelength. Therefore, the directivity of the reflector, considering the random phase errors [see Eq. (11.23)], may be written as

$$D = (4\pi S/\lambda^2)\ g \exp(-\overline{\Phi}^2) = (\pi D_r/\lambda)^2\ g \exp[-(4\pi\overline{\Delta}/\lambda)^2] \qquad (14.11)$$

where D_r = reflector diameter

g = resultant area efficiency factor subject to the effect of all other factors (except the random errors) that impair the directivity of the antenna

$$\exp(-\overline{\Phi}^2) = K_\sigma$$

It follows from Eq. (14.11) that the directivity of a parabolic reflector antenna in the direction of maximum radiation increases inversely as the wavelength squared only in the case of negligible errors in the reflector contour when one may neglect the last exponential factor with a negative power index. If the reflector diameter D and the rms contour error $\overline{\Delta}$ are fixed in advance, a decrease in wavelength will gradually slow down the rise in directivity and will finally cause it to decrease abruptly.

The directivity as given by Eq. (14.11) is a maximum

$$D_{max} = 0.023g\,(D_r/\overline{\Delta})^2$$

at a wavelength

$$\lambda_{min} = 4\pi\bar{\Delta}$$

and at $K_\sigma = -4.34$ dB. At shorter wavelengths, K_σ rapidly decreases, and the directivity goes down. Thus random errors in the surface contour of a reflector limit the directivity of a parabolic reflector antenna and set the limit for the lowest wavelength λ_{min} that may be used. It may also be noted that the limit for the largest wavelength for a parabolic reflector antenna may be set at $\lambda_{max} = D_r/3$ which corresponds to the limit of applicability for physical optics concepts in the calculation of reflector radiation by the induced current method.

From studies into the manufacture of parabolic reflectors by various methods, it has been found that the ratio of the rms error in the surface contour to the reflector diameter, or the relative error, is a nearly constant quantity, $\bar{\Delta}/D_r \approx$ const, peculiar to a particular manufacturing technique. For the most common methods of reflector manufacture in quantity

$$\bar{\Delta}/D_r \approx 0.4 \times 10^{-3}$$

which means that for a reflector 1 m in diameter the tolerance on the surface contour may be

$$\Delta_{max} \approx 1 \text{ mm}$$

In order to obtain

$$\bar{\Delta}/D_r \leqslant 0.4 \times 10^{-4}$$

a finer reflector technology is essential, often involving surface fitting. At this writing, one of the world's most precise units is the unique parabolic reflector with a diameter of 22 m, built for the Soviet radio telescope RT-22 near the city of Serpuhov in the Soviet Union in 1958. The tolerance on the surface contour of this reflector is 0.5 mm and the relative accuracy is

$$\bar{\Delta}/D_r = 0.9 \times 10^{-5}$$

The correlation radius for surface discontinuities is about 20 cm which is equivalent to from two to twenty-five wavelengths in the operating wavelength band from 10 cm to 8 mm. At a wavelength of 8 mm, the beamwidth of the RT-22 radio telescope is 2 minutes of arc, the effective aperture area being 150 m^2 (with an overall aperture efficiency of around 0.45).

Very precise parabolic reflectors of a smaller size can be fabricated by spinning a molten plastic material until it cools and solidifies. The parabolic surface thus obtained is then metallized, and its relative accuracy is

$$\bar{\Delta}/D_r = 2.4 \times 10^{-5}$$

for a diameter of 8.5 m.

For the purpose of comparing circular parabolic antenna of different diameters in terms of directivity, Fig. 14.17 shows plots of directivity as a function of operating frequency for various reflector

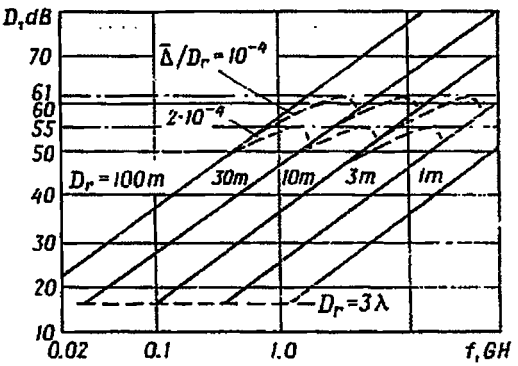


Fig. 14.17. Directivity of a reflector antenna as a function of frequency

size tolerances and for $g = 1$. Recall that a reduction in directivity due to random phase errors occurs because of an increase in the uniform sidelobe radiation while the shape of the main radiation beam remains nearly undistorted.

14.5. Feeds for Reflector Antennas

The feed is a key component of a reflector antenna as it has a decsive bearing on the performance of the entire antenna system. Ty pically, a feed is to satisfy the following requirements.

1. The field-strength radiation pattern should give the desired amplitude distribution in the aperture and produce a minimal radiation outside the aperture angle.
2. The directional phase characteristic should have a phase centre located at the focal point of the reflector.
3. The cross-sectional dimensions of the feed must be such as to minimize aperture blocking.
4. The electric strength of the feed should be high enough to handle the full operating power of the radio system without any risk of breakdown.
5. The bandwidth of the feed should be compatible with that of the radio system.
6. The physical structure of the feed must be rugged enough to stand up to changes in the weather conditions and it must allow the microwave transmission system to be hermetically sealed. As a rule, the feed will be placed inside a dielectric enclosure.

Consider some of the feed types for parabolic reflector antennas.

Dipole feeds. As a rule, a dipole feed consists of a driven dipole and a subreflector in the form of a disc or a parasitically excited reflector. The driven dipole can be energized by a coaxial line or a rectangular waveguide. Two classical designs of a dipole feed energized by a coaxial line ($Z_w = 50$ ohms) and fitted with a disc subreflector are shown in Fig. 14.18. The first design uses a balun in the form of a quarter-wave skirt, whereas the second design uses a slot-balancing stub. The latter device provides a better balancing

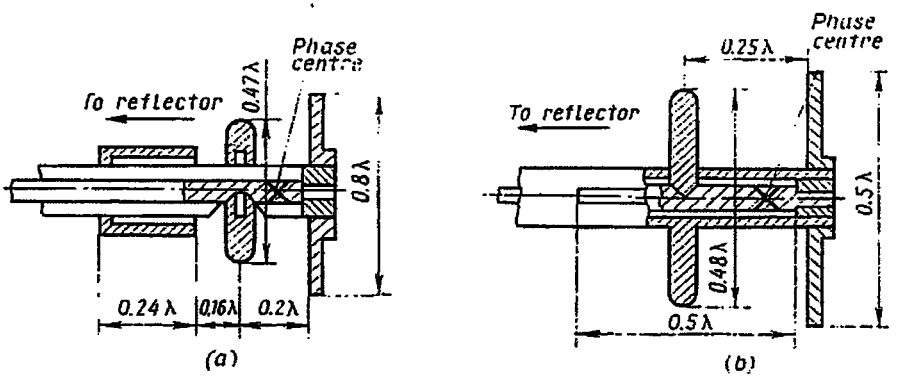


Fig. 14.18. Coaxial-fed dipole feeds: *a*, $A = 0.47\lambda_0$, $B = 0.19\lambda_0$, $C = 0.80\lambda_0$, $D = 0.24\lambda_0$, $E = 0.16\lambda_0$; *b*, $A = 0.48\lambda_0$, $B = 0.25\lambda_0$, $C = D = 0.5\lambda_0$; × = = phase centre

effect, but its power-handling capacity is lower as it is limited by the allowable voltage across the slot. In each case the phase centre is located between the dipole and the subreflector, the optimum aperture angle of the reflector being 140 to 160°.

In cases where it is important to have a capability for simultaneous operation with two orthogonal polarizations, use may be made of turnstile dipoles in which each pair of arms is energized by a separate coaxial line, with the two lines laid in a common enclosure.

The feeding method for a dipole feed by a rectangular waveguide is shown in Fig. 14.19. The dipole around 0.7λ long is secured in a PTFE sleeve which is in turn placed inside a TE_{10} rectangular waveguide of dimensions $0.15 \times 0.7\lambda$, tapered in the narrow dimension. Instead of a subreflector, use is made of a shorted coaxial line whose inner conductor is the waveguide. In terms of radiation characteristics, this combination of a driven dipole and a passive radiator in the form of the aperture of a shorted coaxial line operating in the TE_{11} mode is close to a Huygens source with a cardioid radiation pattern. The beam width is the same in the *E* and *H* planes and equal to about 130° between 10-dB points, with the backward ra-

diation attenuated by 30 dB. The feed provides for an area efficiency of around 0.65 and can operate over a bandwidth of around 12% with an SWR of not over 1.5. This type of feed can be adapted for operation with circular polarization. In such a case, the waveguide should be circular in cross-section and loaded with a dielectric, and the dipole should be a turnstile type.

The dual-aperture (Cutler) feed. This type of feed is convenient to use at low levels of radiation in the SHF band. In sketch form

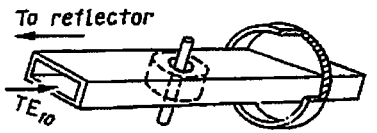


Fig. 14.19. Waveguide-fed dipole feed. Waveguide dimensions 5.1×22.9 mm. Dipole length 22.9 mm. Resonator diameter 36.8 mm. Resonator depth 8.9 mm. Frequency band 8.5-9.6 GHz

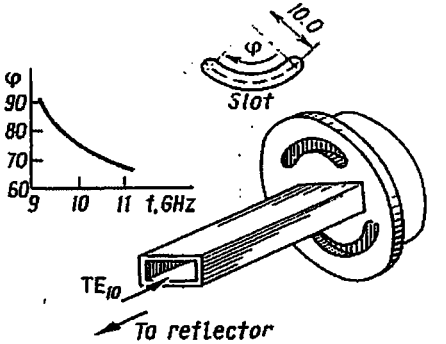


Fig. 14.20. Dual-aperture (Cutler) feed. Waveguide dimensions 5.1×22.5 mm. Resonator length 22.2 mm. Resonator depth 4.9 mm. Flange diameter 44.5 mm

the Cutler feed is shown in Fig. 14.20. The curvature and relative position of the two apertures cut in the end of a cylindrical resonant cavity, nearest to the reflector, and the diameter of the circular shield are chosen so as to obtain the same radiation pattern width (around 130°) in the E and H planes as measured between —10 dB points. For a proper match of the feed to the guide, the latter is tapered in the narrow dimension and a tuning screw is provided, which is screwed into the wall of the resonant cavity, farthest from the reflector. With this type of feed, the aperture area efficiency of the reflector is around 0.57 (considering the spillover). Unfortunately, the dual-aperture (Cutler) feed has a limited bandwidth (about 3%).

Waveguide horn feeds. These are simple in design, have a broad bandwidth, and offer an opportunity to obtain a desired radiation pattern between broad limits. As a rule, waveguide horn feeds (Fig. 14.21) exciting a dominant TE_{10} mode in a rectangular guide and a TE_{11} mode in a circular guide are used relatively seldom now because they do not provide axial symmetry for the radiation pattern whereas the bending of electric and magnetic lines of force in their aperture gives rise to an appreciable cross-polarization of the radiated field.

At this writing, by far the best types are waveguide horn feeds with a corrugated inside surface (Fig. 14.22). The annular corrugations are close to a quarter-wavelength deep and are spaced not more than 0.1λ apart. At the interface between the corrugated-surface structure and the inner space of the horn the boundary conditions for

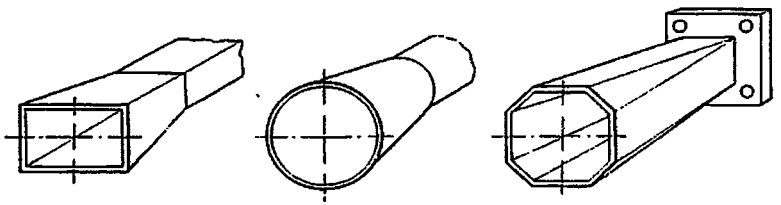


Fig. 14.21. Waveguide horn feeds

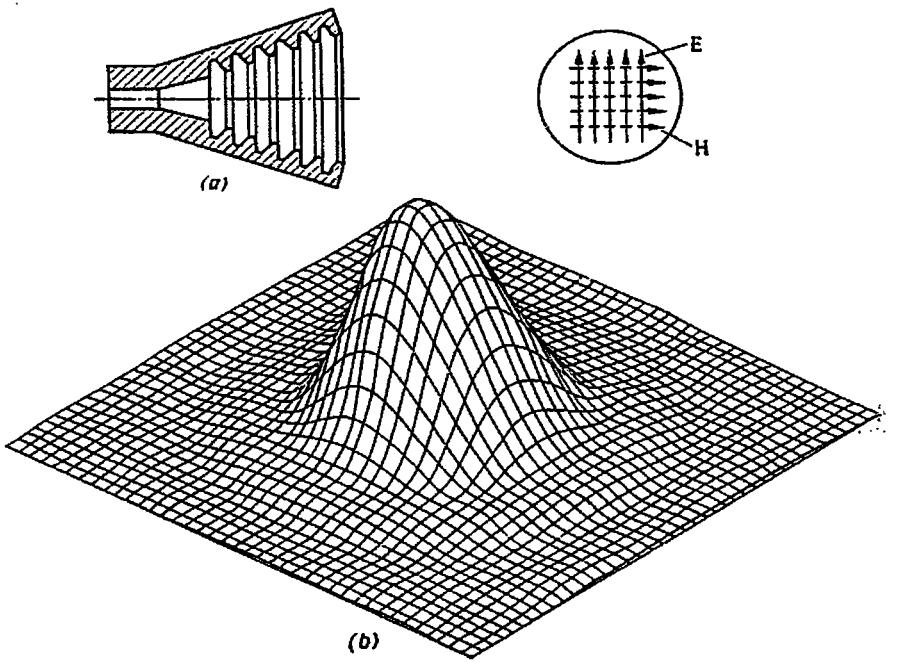


Fig. 14.22. Corrugated horn feed: *a*, sketch; *b*, radiation pattern

the azimuthal components of the E and H fields are nearly the same — a factor which permits the existence of a hybrid HE_{11} wave in which the electric and magnetic lines of force are bent only slightly, if at all, over the cross-section. As a result, the radiation pattern possesses axial symmetry and there is practically no cross-polarization. The depth of corrugations (if it is not close to a half-

wavelength) and corrugation spacing affect only slightly the shape of the radiation pattern of a corrugated-surface horn. The frequency behaviour of a corrugated-surface feed is likewise only slightly dependent on the inside diameter $2a$: at $a/\lambda \geqslant 1.2$, the phase velocity of the HE_{11} wave changes but little with frequency, and approaches that of light. For this reason, a corrugated-surface horn should have a rather large aperture. Calculations and experiments show that with the operating frequency changing by as much as 50% the radiation pattern width of a corrugated horn feed will change by as little as 3% to 7% (depending on the aperture angle and horn length).

The parameters of a corrugated horn feed can further be optimized by placing it in the aperture of a suitably contoured dielectric lens. The lens will both seal the feed and raise the area efficiency of the reflector antenna to nearly 0.9. When use is made of hybrid modes with an odd distribution of the transverse electromagnetic field, corrugated feeds offer a means for producing difference radiation patterns with a radiation null in the direction of the reflector axis.

Line-source feeds. Parabolic-cylinder reflector antennas (see Fig. 14.12*b*) need feeds in the form of linear antennas of a length equal to that of the directrix of the reflector. In the xOz plane, the radiation pattern of such feeds should provide the desired amplitude distribution in the aperture of the parabolic cylinder. Also, the current distribution in amplitude and phase along the y-axis in the feed aperture must be such as to ensure the desired shape for the radiation pattern of the entire antenna in the yOz plane. Linear feeds may take the form of multislot waveguide arrays, sectoral horns, microstrip linear arrays, etc. As often as not, a linear feed is made in the form of a *cheese* or *pillbox reflector* (see Fig. 14.12*c*) which is in turn excited by the open end of a rectangular waveguide. The spacing Δ between the plates of a pillbox reflector is chosen such that higher modes cannot be propagated. If the E vector is normal to the plates, the dimension Δ may be smaller than a half-wavelength. In such a case, the velocity of propagation for TEM waves between the plates will be equal to that of light. If, on the other hand, the E vector is parallel to the plates, the dimension Δ should be chosen anywhere between a half-wavelength and a wavelength, and the TE mode will be propagated between the plates. The tolerance on the dimension Δ is tighter in the latter case.

Reflector-antenna beam position control by an off-focus and an off-axis feed. In order that the beam of a reflector antenna could be slewed or nutated within a small sector while the reflector remains stationary, the phase centre of the feed, *2*, may be offset a short distance Δl away from the focal point, *1* (Fig. 14.23). The rays reflected from the paraboloid will be then deflected away from the axis of the antenna in the opposite direction and make with it

an angle α which can be found from the formula

$$\tan \alpha \approx \Delta l/f$$

The wave front in the reflector aperture and the direction of maximum radiation will also be slewed through the same angle α. At the same time, however, this gives rise to additional cubic phase errors (coma) in the aperture which cause the beam to broaden and the sidelobe level to rise (see Fig. 11.11). So long as the distance Δl is small, these errors are insignificant. Therefore, the transverse offset of the feed may be utilized to nutate the beam within twice or three times its width.

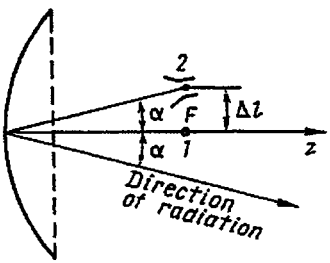


Fig. 14.23. Off-focus feed

The offset of the beam in a parabolic reflector antenna from the reflector axis is utilized in radars for automatic tracking in azimuth and elevation. Two methods are mainly used, conical scanning and monopulse. With *conical scanning*, the antenna beam is rotated about the reflector axis or the *boresight* at a rate of several hertz by rotating an off-axis feed. This rotation produces an equisignal direction on the boresight. When the target of interest is off the boresight, or the equisignal direction, the return signal is amplitude-modulated at the rotation frequency of the feed. The magnitude and phase of the modulation determine the magnitude and direction of the target's motion, and an error signal is generated that actuates the autotracking system of the radar set. The sensitivity in azimuth and elevation is a maximum when the beam is offset from the reflector axis by about a half of its width (the gain along the reflector axis is decreased by about 3 dB).

Monopulse feeds. In the simplest monopulse system intended to locate targets in only one plane (say, in azimuth), use is made of a composite feed which consists of two horns energized by a magic-T or by a hybrid ring, as shown in Fig. 14.24*a*. When the Σ-input is energized from the H-arm of the magic-T, the two horns of the feed are excited in phase and in equal magnitude, thus forming in space a sum radiation pattern whose peak is oriented along the boresight (Fig. 14.24*b*). When the Δ-input is excited from the E-arm of the magic-T, the two horns are driven in anti-phase and produce in space a difference radiation pattern (see Sec. 13.4). The target of interest is illuminated only by the sum pattern, whereas the echo signals are picked up by two receivers which collect both the difference and the sum beams. The received echos are compared in terms of amplitude and phase at the outputs of the intermediate-frequency amplifiers of the receivers, and the comparison gives both the magnitude and direction of the target's motion off boresight (the antenna

axis). In principle, it would suffice to pick up only one echo pulse to accomplish the task. Hence, the name 'monopulse system'.

If the difference pattern is to have an optimum shape, the phase centre of each horn should be offset from the reflector axis by the same amount as would be necessary in a conical-scan system. When

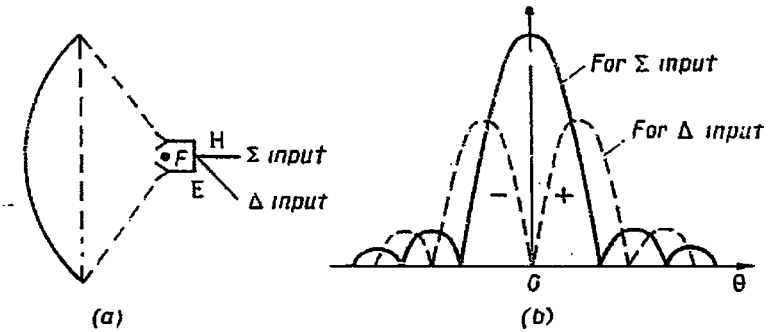


Fig. 14.24. To operation of a monopulse feed

the position of a target is to be determined in both azimuth and elevation, the monopulse feed should consist of at least four horns excited via a suitably arranged beam-shaping network containing a certain set of hybrid junctions.

14.6. Special Types of Reflector Antennas

The Cassegrain two-reflector antenna. In a Cassegrain (or Cassegrainian) antenna, an in-phase field distribution in the aperture is produced by two reflecting surfaces. One of them is the main (large) paraboloidal reflector, *1*, and the other is a hyperboloidal (small) subreflector, *2* (Fig. 14.25). One of the foci, F_1, of the hyperboloidal subreflector is placed at the focal point of the main paraboloidal reflector, whereas the phase centre of the primary feed, *3*, is located at the other focus, F_2, so that the rays reflected from the subreflector may be regarded as emanating from a 'virtual feed', *4*, located behind it at point F_1. The image of the virtual feed is formed by the rules of geometrical optics and is scaled down in size.

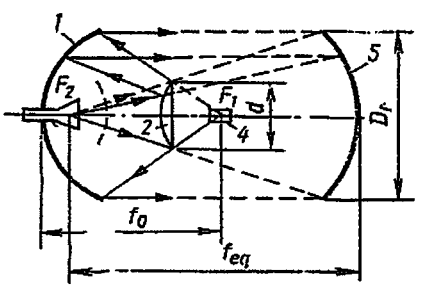


Fig. 14.25. Cassegrain two-reflector antenna

In determining the amplitude distribution it is convenient to replace the two-reflector system by an equivalent paraboloid, 5,

of the same diameter as the main reflector, but with a greater focal length f_{eq}. Construction of the equivalent paraboloid will be clear from Fig. 14.25. For a given primary feed, there is an optimum ratio f_{eq}/D_r at which the overall area efficiency, or the product of $K_{ae,a}$ and $K_{ae,f}$ is a maximum (similarly to a single-reflector antenna, see Fig. 14.15). The aperture of the main reflector in a Cassegrain

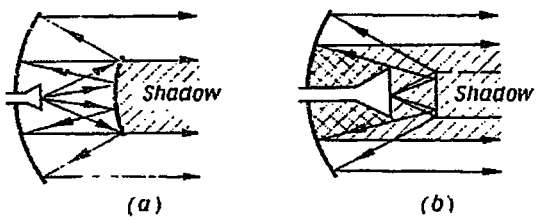


Fig. 14.26. Aperture blocking in a Cassegrain two-reflector antenna: *a*, blocking mostly by the subreflector; *b*, blocking mostly by the feed

antenna is blocked by both the subreflector and the primary feed (Fig. 14.26). Aperture blocking is a minimum when the shadow areas produced by the primary feed and by the subreflector are equal in size.

The Cassegrain two-reflector antenna compares favourably with a single-reflector system on a number of counts. For one thing, it offers an opportunity to use a shorter microwave transmission system and to locate the larger part of the primary feed structure behind the reflector, which is especially convenient in monopulse radars. When the dimensions of the primary feed and of the subreflector are properly optimized, a relatively high overall area efficiency (0.60 to 0.65) can be obtained. The reduced focal length of the main reflector and the low scatter of the feed field into the backward hemisphere of the main reflector serve to reduce the noise temperature. Among the demerits of the Cassegrain system are the back reaction of the subreflector to the primary feed and an increased blockage of the aperture as compared with a single-reflector antenna.

The Gregorian two-reflector antenna. This type of antenna (Fig. 14.27) likewise uses two reflecting surfaces one of which is a paraboloidal (large) reflector and the other is an elliptical reflector. One of the foci, F_1, of the elliptical reflector is located at the focal point of the main paraboloidal reflector, whereas the other focus, F_2, is located at the phase centre of the primary feed. Waves emanating from the primary feed arrive at the elliptical reflector which focuses them at point F_1 and re-directs them onto the main paraboloidal reflector, so that an in-phase wave front is produced in its aperture. As theoretical calculations and experiments show, the overall aperture a efficiency of a Gregorian two-reflector system is

somewhat higher than that of a Cassegrain system. The Cregorian system is preferable for large, all-round-looking antennas.

Quasi-paraboloidal two-reflector antennas. The classical paraboloidal and elliptical reflectors of the Gregorian system can be modified in such a way that the amplitude distribution is made more uniform, thus enhancing the overall aperture efficiency while retaining the in-phase field distribution in the aperture of the main reflector. The ray paths in such an optimized antenna with a 'quasi-paraboloidal' main reflector are shown in Fig. 14.28. Note that

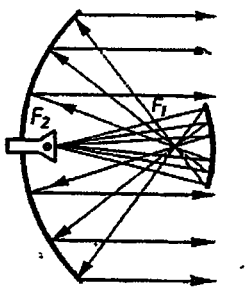


Fig. 14.27. Gregorian two-reflector antenna

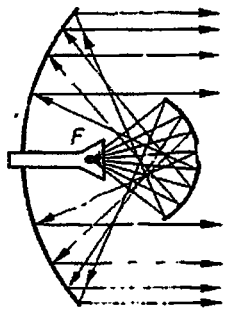


Fig. 14.28. Quasi-paraboloidal two-reflector antenna

the rays reflected from the elliptical reflector have no common point of intersection and that the "crowding" of rays towards the aperture edges, causing a rise in the amplitude distribution at the edges.

The idea of a quasi-paraboloidal antenna, first proposed by L.D. Bakhrakh of the Soviet Union, has been favourably received in the world and embodied in many major radio telescopes. Notably, this idea has been implemented in the RT-70 all-round-looking antenna with a main reflector diameter of 70 m built near Evpatoria in the Crimea, USSR, for deep-space communications in 1978. It has the highest aperture area efficiency of all the major radio telescopes in the world, the figure being 0.80 in the UHF band and 0.74 in the SHF band. As in other major Soviet radio telescopes, the RT-70 utilizes the so-called homologous distribution of reflector strain as it is tilted in elevation. The stiffness of the load-bearing reflector frame is distributed so that, as it is strained in tilting, the reflector acquires the shape of a distorted quasi-paraboloid slightly turned in space, with an off-axis focal point. By moving the primary feed to the new position of the focus with the aid of an automatic control system, it is possible to retain a high overall aperture area efficiency as the reflector is tilted in elevation. The

idea of a homologous strain distribution has been advanced independently by P.A. Kalachev of the Soviet Union and von Hörner of West Germany. The RT-70 also includes a system which effects the phase compensation of gravitational strain, owing to which this mammoth antenna can be used at wavelengths down to 8 mm with a reduction of a mere 20% in the antenna aperture efficiency as the refector is tilted from zenith to the horizon.

Off-set and hybrid reflector antennas. In axi-symmetric single- and two-reflector antennas, the primary feed is located in

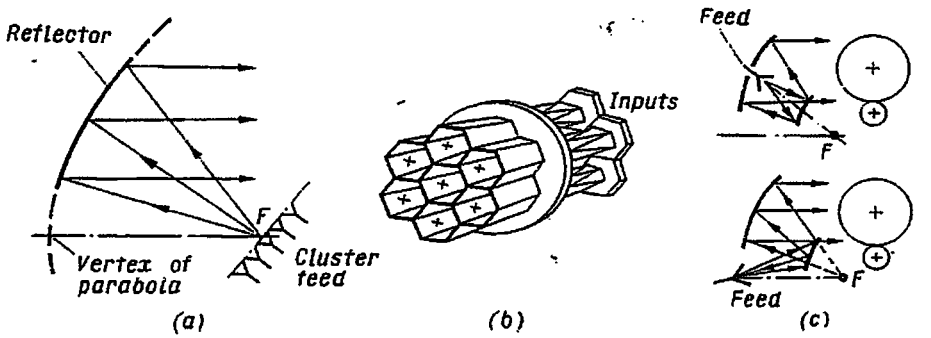


Fig. 14.29. Off-set reflector antennas: *a*, single-reflector [hybrid antenna; *b*, segment of a multielement feed; *c*, off-set Cassegrain two-reflector antennas

the field of the rays reflected from the main reflector or the subreflector. This impairs the aperture efficiency of the reflector and brings on a rise in the side lobes of the radiation pattern due to aperture blocking and backward radiation towards the primary feed. The latter event consists in that the reflected rays intercepted by the primary feed mismatch the input, and there is nearly no way of compensating for this mismatch by ordinary matching devices because the primary feed is located a relatively long distance away from the reflector.

The above drawbacks are all but eliminated in *off-set* or *off-axis reflector antennas* (Fig. 14.29) in which the feed system is placed outside the field of the rays reflected from the main reflector. Unfortunately, the reflector in such cases is the shape of an asymmetrical section from a rotational paraboloid which is difficult to fabricate, and the overall dimensions of the antenna greatly exceed the size of the radiating aperture. Since there is no aperture blocking by the feed, the primary feed for an off-set reflector antenna may be in the form of an array (or cluster) of closely spaced elements (Fig. 14.29*b*). The input of each array element corresponds to a separate radiation pattern. The radiation patterns of all the

elements closely fill a relatively small sector of solid angles within approximately plus or minus 5° of the major axis of the paraboloid (with greater angles of tilt from the axis of the paraboloidal reflector, it becomes increasingly more difficult for the deflected rays to retain a regular shape because of systematic phase errors). By switching the inputs of the array elements or by redistributing among them the radiated power in a certain amplitude and phase relationship, it is possible to control the direction of the beam or to control the shape of the resultant radiation pattern in exactly the same way as this is done by the partial radiation-pattern method in antenna synthesis theory.

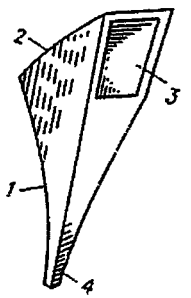


Fig. 14.30. Horn parabolic reflector antenna

A combination of one or several reflecting surfaces (or lenses) with a multielement cluster feed is usually referred to as a *hybrid antenna system* or, simply, a *hybrid antenna*. A hybrid antenna combines the advantages inherent in its component types and avoids their drawbacks. There is a growing tendency to install hybrid reflector antennas on geostationary communications satellites (as seen from a geostationary orbit, the Earth's diameter subtends an angle of around 9°). Such hybrid reflector antennas offer a means for establishing narrowband channels between communicating parties spaced great distances apart on the ground or, when as many as tens of channels are used, for distributing the radiation among the desired regions on the Earth's surface. This is the contour-beam radiation pattern which takes the shape of the respective country or a part of it. As an alternative, off-set and hybrid antennas can be of multireflector types (Fig. 14.29*c*).

The horn parabolic reflector antenna. This type of antenna (Fig. 14.30) is a combination of a pyramidal horn, *1*, and a segment of a rotational paraboloid, *2*. The focus of the parabolic surface is placed at the vertex, *4*, of the horn. In consequence, the electromagnetic waves emanating from the horn and illuminating the parabolic reflector are reflected towards the aperture, *3*, as a parallel beam of rays, without going back to the horn. Horn parabolic reflector antennas have a low sidelobe and backward radiation level unattainable by other types of reflector antennas. This makes them ideally suited for use in microwave radio-relay links. The sidelobe level at the aperture plane is around 50 dB down, and the backward radiation level is around 60 dB down. The flare angle of the horn is usually made to be anywhere between 30 and 40°. At the horn input there is a tapered matching transition from seven to ten wavelengths long. This brings down the reflection coefficient to 1 or 2% over a very broad bandwidth, for which reason a horn para-

bolic reflector antenna can operate in two (or even more) radio-relay bands at a time.

Periscope two-reflector antennas. This type of antenna (Fig. 14.31) finds use in microwave radio-relay links. It consists of an upper reflector, *1*, a lower reflector, *2*, and a primary feed. The upper reflector is usually made flat, with an elliptical or polygonal contour so that its projection on a plane normal to the direction of radiation is a circle or a polygon. The lower reflector is given a special concave shape such that an optimum current distribution in amplitude

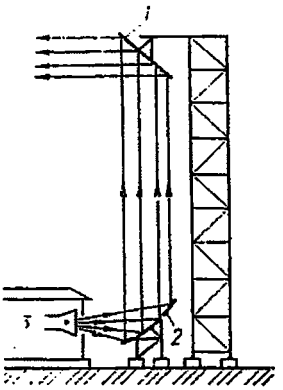


Fig. 14.31. Periscope two-reflector antenna

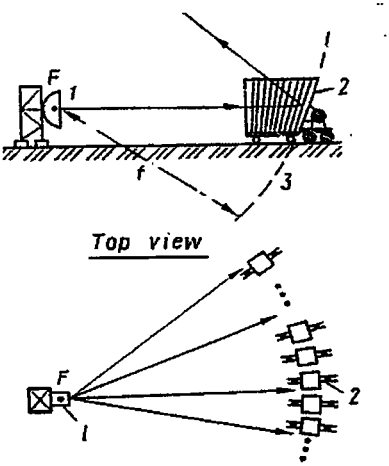


Fig. 14.32. Multiplate antenna

and phase is produced in the equivalent aperture of the upper reflector. The surface of the lower reflector is close to that of a section from an ellipsoid of revolution with one of its foci located at the centre of the horn and the other at the point of suspension of the upper reflector. The primary horn feed, *3*, is located in a nearby equipment building. A major advantage of a periscope two-reflector antenna is that there is no need for a long microwave transmission system; its major limitation is poor noise immunity because a good deal of the electromagnetic energy supplied by the horn feed is scattered on its way first to the lower reflector and then to the upper one.

The multiplate antenna. S.Ye. Haikin and N.L. Kaidanovsky of the Soviet Union have proposed an antenna system close in the basic idea to the periscope two-reflector type, intended for use as a high-resolution radio telescope (Fig. 14.32). It has a stationary primary feed, *1*, positioned at the focal point of a reflector, *2*, assembled from a great number individually adjustable metal reflect-

ing plates. Each plate can be moved radially and turned about its vertical and horizontal axes. In shape, this composite reflector is a section of a rotational paraboloid, *3*, and, as such, it converts the spherical wave emanating from the primary feed at its focus to a plane electromagnetic wave at its aperture. The position of the beam in space can be controlled by varying the tilt and position of the plates.

An important thing about this multiplate antenna is the fact that the contour accuracy of the equivalent reflector is determined by the positioning accuracy of each plate. Therefore, the equivalent reflector can be made so that the relative contour accuracy, $\overline{\Delta}/L$, is about one part in a million. Here, $\overline{\Delta}$ is the rms contour error and L is the dimension of the equivalent aperture. This level of contour accuracy is unattainable with fully steerable parabolic antennas.

The multiplate antenna concept has been embodied in RATAN-600, the largest radio telescope in the USSR at this writing. It consists of 900 reflecting panels each measuring 7.4 m by 2 m, installed all the way round a circle 588 m in diameter. Each reflecting panel can be adjusted for position in such a way that the incident radiation from the radio source under observation is re-directed to the focal point F. This idea of collecting electromagnetic energy goes back to Archimedes who, according to a legend, proposed to burn the enemy's fleet by beaming sun light reflected by a great number of flat mirrors onto the ships. In the RATAN-600, the beamwidth is 4.3 seconds of arc at the lowest operating wavelength of 8 mm. The noise temperature does not exceed 30 K. Each panel is positioned as required by a computer.

Spherical reflector antennas. With a primary feed positioned at a point $f = R/2$, a spherical reflector is similar in performance to a parabolic reflector (Fig. 14.33). If we compare the equation of a circle

$$x^2 + (R - z)^2 = R^2$$

or

$$x^2 = 2Rz\,(1 - z/2R)$$

with the equation of a parabola

$$x^2 = 4fz$$

we may conclude that the part of the circle near the reflector vertex 0 (with $z/2R$ substantially smaller than unity) may be regarded as a segment of a parabola whose focus is located at a point $z = R/2$. A spherical reflector illuminated from the point $R/2$ will convert a spherical diverging wave to a nearly uniform plane wave, for which reason it may be used to shape sharp radiation patterns. By moving the primary feed on the focal sphere of radius $R/2$, it is possible to

deflect the radiation pattern between broad limits without any deterioration in its shape while the reflector remains stationary.

A stationary spherical reflector can be built as a pit dug in the ground and covered with reflecting panels. Since the panels are firmly secured to the ground, the reflector will not be subject to mechanical strain. With this arrangement, it is possible to build a scanning-beam antenna with a huge reflector to extremely tight tolerances on its contour and at a relatively low cost. This principle is embodied in the ROT-32/54, a unique Soviet radio-optical telescope located on the southern slope of Mount Aragaz, Armenia, at 1800 m above sea level. The stationary main reflector of this antenna, designed by P.M. Geruni of the Soviet Union, has the shape of a hemisphere 54 m in diameter. It is assembled from 3800 panels installed in a concrete dish which is in turn built in a hard-rock pit. Each panel is cast of an aluminium alloy and machined on a rotary lathe so that the surface irregularities are not more than 10 μm high and the panel surface has a radius of curvature of 27 m. The panels are set up on posts 1.8 m tall so that their position can be adjusted to fit the radius of a sphere.

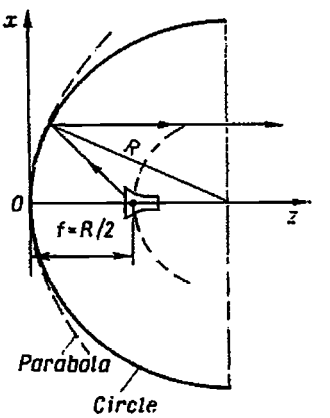


Fig. 14.33. Profile of a spherical reflector

The operating part of the spherical reflector, 32 m in diameter, is illuminated by an auxiliary reflector antenna about 5 m in diameter, set up on three supports so that it can be rotated about two axes around the centre of the large sphere. By moving the feed, it is possible to change the direction of the beam within a 120-degree cone. The automatic control systems of the feed make it possible to steer the beam accurate to within at least 2 seconds of arc. The effective surface area of the antenna is 560 m^2, the noise temperature is 10 K, the beamwidth is 7 seconds of arc, and the gain is 10^9 at a wavelength of 1 mm. The feed counterweight assembly set up over the centre of the sphere gives support to an optical telescope with a reflector 2.6 m in diameter. Thus, in addition to radio-beam observations, astronomers can carry out observations of sources in the optical and infra-red bands. The ROT-32/54 offers a means for observation of not only celestial radio sources but also the Earth's atmosphere, for long-distance space communication, and for the search of extra-terrestrial civilizations.

A spherical reflector may alternatively be used for a circular scan, as shown in Fig. 14.34. The reflector surface, *2*, is formed by rotating a part of a circle about a vertical axis. The polarization-selective reflector is assembled from parallel plates inclined at an angle of

45° and spaced a fraction of a wavelength apart, so that the plates on the opposite sides are at right angles to each other. The primary feed, *1*, is located on the focal circle, *3*, of a half radius, and sets up a field with a linearly polarized E vector inclined at 45° to the horizontal and parallel to the plates. On the opposite side of the reflector, the plane of polarization of the reflected wave is perpendicular to the plates, because of which the wave is free to pass through this part of the reflector almost without losses. Depending on the requirements that the radiation pattern is to meet in the vertical plane, the reflector contour in that plane may be made parabolic or otherwise shaped. Circular scan is obtained by moving the primary feed on the focal circle.

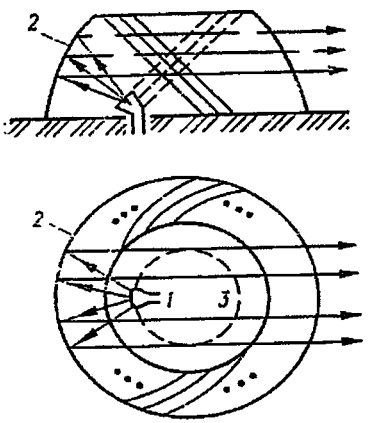


Fig. 14.34. All-round-looking spherical antenna with a polarization-selective reflector

The area efficiency of spherical reflector antennas can be increased through the use of compensating feeds which have a suitably shaped phase characteristic that would eliminate the systematic phase errors arising because the actual contour of a real spherical reflector differs from that of an ideal parabola. Incidentally, this idea is utilized in the ROT-32/54 radio telescope.

Shaped-beam antennas. This term applies to an antenna whose radiation pattern over a certain angular range is designed to a special

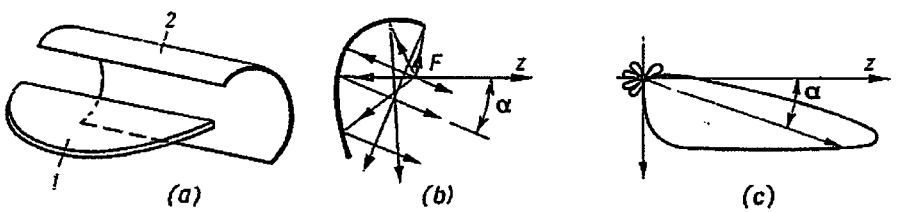


Fig. 14.35. Cosecant-squared antenna

shape for some particular use, notably a cosecant-squared beam. One classical example is the antenna shown in Fig. 14.35*a*. It consists of a specially shaped cylindrical reflector which produces a cosecant radiation pattern in the vertical plane and a linear feed in the form of a cheese (or pillbox) reflector which produces a pencil beam in the horizontal plane. The lower part of the shaped reflector is nearly parabolic and it collimates the reflected rays into a nearly parallel beam. The upper part is nearly a circular cylinder in shape

and reflects the incident rays as a divergent beam (Fig. 14.35*b*). The resultant radiation pattern is shown in Fig. 14.35*c*. Reflector contours for shaped-beam antennas are designed on a computer, using specialized programs for reflector antenna design.

Chapter Fifteen

Arrays

15.1. Scanning Methods. Objectives Served by Arrays

Scanning, or the motion of an antenna beam in space, can be effected mechanically, electromechanically, or electronically. In mechanical scanning, the entire antenna structure is rotated in some way—a fact which heavily limits the scan rate and calls for high power expenditure. In electromechanical scanning effected with the aid of electromagnets or electric motors, one or several elements of an antenna are moved so that the equiphase surface of the radiation field is tilted in a stationary aperture. One classical example is the control of the beam position in a reflector antenna through the use of an off-axis feed. Electromechanical scanning gives a higher scan rate because the moving elements have a smaller mass compared with that of the entire antenna structure. However, neither mechanical nor electromechanical scanning satisfies the state-of-the-art requirements for scan rate, nor can several targets be tracked at a time in either case.

The highest scan rate is obtained with electronic scanning. With this method, the current distribution in a mechanically fixed aperture is varied with the aid of electronically controlled devices, such as semiconductor or ferrite phase shifters and switches. Scan rate here is limited by the inertia due to the time constant of electronic circuits, but this inertia is by several orders of magnitude smaller than the mechanical inertia inavoidable in the first two methods.

Electronic scanning is effected in multielement arrays. It can further be subdivided into *phase scanning*, *amplitude scanning*, and *frequency scanning*. In phase scanning, one solely varies the phase difference between adjacent array elements while the amplitude distribution remains almost unchanged. In amplitude scanning, the beam is scanned by switching the inputs of a multibeam antenna system; this technique is more commonly called *beam-switching*. In frequency scanning, the required phase difference between the array radiators is provided by adjustment of frequency, and this calls for the use of suitably configured frequency-dependent circuits to energize the array elements.

The key principles of beam-position control in arrays by varying

the phase of excitation applied to the array radiators were known as early as the 1920s-30s, but it was not until the 1950s-60s that practical electronically scanned antennas were built. Their development was spurred by a continuous rise in the speed of aircraft, the appearance of missiles, and the early ventures in space exploration, all of them stimulating the demand for high-speed radar facilities. Electronic scanning received a 'big short in the arm' with the advent of computers without which it would have been inconceivable to coordinate the operation of hundreds or even thousands of radiators in an electronically scanned array. Advances in high-speed semiconductor and ferrite control devices have also served to promote electronic scanning.

The increase in the number of elements used in antennas has had as its corollary an increase in the range of tasks that could be tackled by radio systems and has given rise to new principles in radio communication, radar, and radio astronomy. We will name only the most important of them.

1. Electronic scanning over a wide sector.
2. Beam shaping by the amplitude and phase adjustments of the currents energizing the individual radiators in an array.
3. Coherent addition in a single beam of the output power from many microwave generators or power amplifiers in order to obtain higher levels of radiated power unattainable in conventional antennas due to the limited electric strength.
4. Better extraction of information from the electromagnetic waves arriving at an electronically scanned array as a result of parallel processing by sophisticated circuits of the signals picked up by the individual elements of an antenna system.
5. In-phase addition of the signals received by a system of major antennas in order to obtain high effective areas in radio reception, unattainable with conventional antennas due to inaccuracies in antenna manufacture.
6. Improved reliability of radio systems owing to the parallel operation of many elements. Failure of, say, 20% of the array elements would not cause a catastrophic failure of the entire radio system; this would only partially impair its performance.

Several specific difficulties stand in the way of building high-performance electronically scanned arrays. Most important of them are the occurrence of phase errors in the aperture due to inaccuracies in operation of the control devices, step-wise phasing, mismatch and mutual coupling between the elements during scan, additional losses of microwave power in the control devices, the need to limit the bandwidth as a way of avoiding grating lobes, and some other factors. One has also to reckon with the high first cost and maintenance expenses of electronically scanned arrays as they need periodic tests, alignment, and replacement of failed components.

It follows from the foregoing that electronically scanned arrays are complex radio systems, and their development involves resort to many principles and concepts from the theory of antennas and microwave devices and information theory. State-of-the-art arrays come in a wide variety of configurations adapted for use on the ground or on board aircraft or space vehicles, using many elements (from units to tens of hundreds), differing in aperture shape (flat, convex, circular, etc.), and operating at various wavelengths (from short waves to wavelengths in the optical region of the spectrum).

15.2. Phased Arrays

Phased arrays fall in the most commonly used class of electronically scanned arrays that can be built as medium-sized and large scanned antenna systems.

Phased arrays may be classed into active and passive. In an *active* phased array, each element is energized by a separate properly phased generator or power amplifier and is also fitted with TR-ATR switches and stages that effect frequency conversion and signal preamplification. All of these components make up between them the transmit-receive module of an active phased array. The module should have small cross-sectional dimensions (six or seven tenths of the wavelength) so that it could be accommodated within the aperture area alotted to each array element. Very often the transmit-receive modules of active phased arrays are built from integrated circuits, microstrip lines, and microstrip radiators. The unquestionable advantages of active phased arrays are adaptability to streamlined manufacture, reliability owing to the use of a multiplicity of elements, and a drastic reduction in the length of the microwave transmission system between the radiators and receivers-transmitters.

In a *passive* phased array, all the radiators are energized from a common generator (or operate into a common receiver). For this reason, it has as an integral part some form of device that divides the input power among the array elements.

Power dividers of the guided type. In this type of power divider, the microwave power is distributed among the array radiators with the aid of passive multiports consisting of transmission-line sections, tees, directional couplers, etc. Such power dividers may energize the array radiators in series or in parallel.

The classical arrangement of a *series feeder system* for a linear equally spaced (or equidistant) array is shown in Fig. 15.1*a*. Power for each radiator is tapped off the main microwave transmission system, and identical feed-through phase shifters are placed in the main transmission system between the taps to adjacent radiators. The tapping devices may be reactive tees with a loosely coupled side

arm, or likewise loosely coupled directional couplers (in which the decoupled arm is terminated in a matched load). This arrangement is compact, and all the phase shifters are controlled in a similar fashion because for the beam to be deflected through a certain angle the phase difference between adjacent radiators must be the same along the length of the array. To state this a bit differently, the rate of phase advance (known as phase taper) should be uniform. This serves to simplify control of the phase shifters. Unfortunately, the series feeder system is subject to a number of drawbacks. Firstly, phase errors and losses tend to accumulate and build up towards the

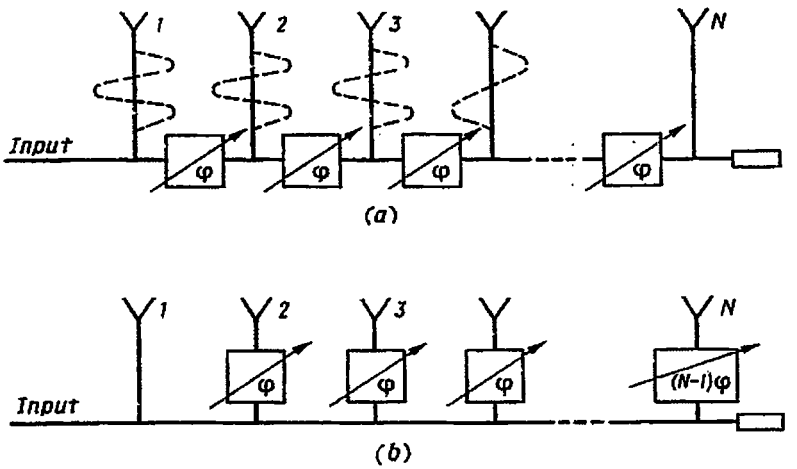


Fig. 15.1. Series feeder system for the radiators of a phased array

end of the array, and this calls for the use of only very accurate, low-loss phase shifters. Secondly, the phase shifter nearest to the input has to handle all of the radiated power and it must be therefore very strong electrically. Thirdly, the electrical length of the path travelled by signals from the common input to each radiator is substantially different, and this might cause a marked de-phasing of the array at the edges of the operating frequency band. This difference in electrical path length can be removed by placing compensating line sections in the feed lines of the radiators (as shown by the dashed lines in Fig. 15.1). Unfortunately, this adds to the overall size of the power divider and it is no longer compact. The phase shifters in a series feeder system may be placed in the side taps from the main transmission line (as shown in Fig. 15.1*b*), but phase-shifter control is then anything but simple.

Figure 15.2 shows a *parallel feeder system* for an N-element array. This arrangement offers a number of advantages. Firstly, relatively

low-power phase shifters may be used because each has to handle one-Nth of the total radiated power. Secondly, the overall power loss in the control devices is solely determined by the attenuation due to only one phase shifter, and this allows the use of phase shifters with an attenuation of 1.0 to 1.5 dB. Important advantages of the parallel feeder system is that phase errors do not accumulate along the aperture and the electrical length of the individual channels can be equalized for a broader bandwidth. A disadvantage of the parallel feeder system is complicated control because the phase taper

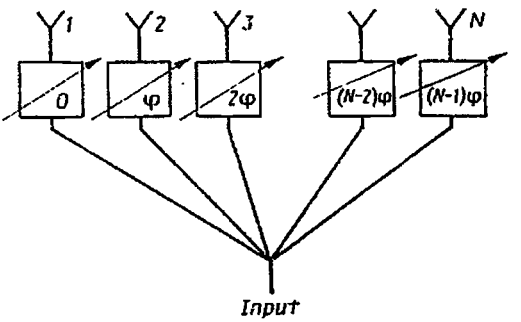


Fig. 15.2. Parallel feeder system for the radiators of a phased array

changes from one phase shifter to the next. Also, some difficulties arise in providing a good match at the power divider input when power has to be split among many channels.

A special case of the parallel feeder system is shown in Fig. 15.3*a*. This is a binary-tier 'herringbone' arrangement in which the input power is consecutively divided by two at each node (other arrangements can be built in which the power is divided into a greater number of parts). The power dividers placed at the nodes of the 'herringbone' structure may be simple tees, hybrids, or resistive ring power dividers. An advantage of this arrangement is the fact that the electrical length is the same for all channels. Unfortunately, this structure is too unwieldy.

The 'herringbone' arrangement can be used in order to simplify array control, if the phase shifters are placed in each tier, as shown in Fig. 15.3*b*. In this case, in order to obtain a uniform phase taper for the radiators spaced 2φ apart, all the phase shifters in each tier should give the same phase advance φ (accurate to the sign). As a result, the number of control signals is reduced to that of tiers in the system.

Planar electronically scanned arrays may use various combinations of series and parallel feeder systems, especially if the entire array is divided into sub-arrays of a smaller size. The choice of a

particular type of feeder system must be based on trade-offs between the allowable attenuation in the control devices, transmitted power, phasing accuracy, and the allowable complexity of the control system. An unquestionable advantage of the above types of feeder systems is the fact that power can uniformly be distributed among the radiators (in fact, any other amplitude distribution can be obtained). Also, there is no uncontrolled parasitic radiation inherent in unguided, or optical, feeder systems.

Unguided (optical-type) power feeder systems. The two basic configurations here are the feed-through type and the reflecting

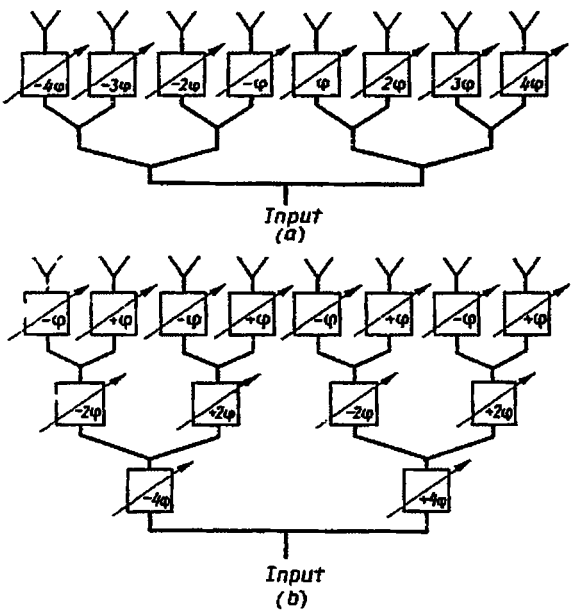


Fig. 15.3. Binary tier feeder system for the radiators of a phased array

type. In a phased array of the *feed-through type* (Fig. 15.4), a primary feed directs the radiated power to a collecting array of receiving elements. The received power passes through a system of phase shifters and, after it has been phased as required, it is re-radiated in the desired direction by another array of radiating elements. Transmission-line sections (shown by the dashed lines in Fig. 15.4) may sometimes be placed between the receiving elements and the phase shifters in order to equalize the electrical length of the paths travelled by signals to the various elements of the radiating array. These line sections may also be used to produce a nonlinear initial phase taper in order to suppress grating

lobes in discrete phasing (see Sec. 15.3). In the principle of operation, a feed-through phased array is equivalent to a waveguide lens within electronically controlled phase distribution in the aperture.

A *reflecting-type phased array* (Fig. 15.5) consists of a primary feed and a receiving-transmitting array in which each element is fitted with a reflecting phase shifter. Delay lines may additionally be placed between the radiators and the phase shifters in order to equalize the electrical length of the paths travelled by signals that

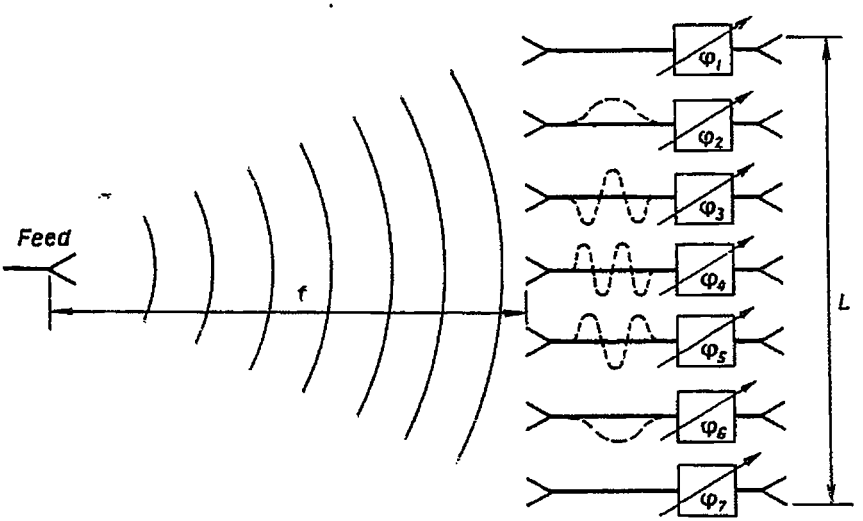


Fig. 15.4. Feed-through optical-type feeder system for the radiators of a phased array

pass through the different array elements and also to produce the required nonuniform initial phase taper. In a reflecting phase array, the radiators do a dual job. Firstly, they collect the power emanating from the primary feed. Secondly, they re-radiate it in the desired direction after phasing. In the principle of operation, a reflecting phased array is equivalent to a reflector antenna with the electronically controlled phase of the reflection coefficient from the various parts of the reflector area.

On the merit side, both types of unguided (optical) feeder systems are relatively simple in design although they consist of a great number of elements, offer convenient control of the amplitude distribution in the aperture by adjustment of the radiation pattern of the primary feed, and permit the use of composite monopulse feeds to produce sum and difference patterns in radars for automatic target tracking. A drawback common to all unguided (optical) feeder systems is an increased size as compared with the guided

types because the ratio of the focal length f to the aperture dimension L ranges anywhere between 0.5 and 1. Also, some of the primary feed power in unguided feeder systems is not intercepted by the receiving array, and this leads to a rise in the sidelobe level and a reduction in the overall aperture efficiency of the antenna. This unpleasant thing can be avoided in a feed-through phased array by

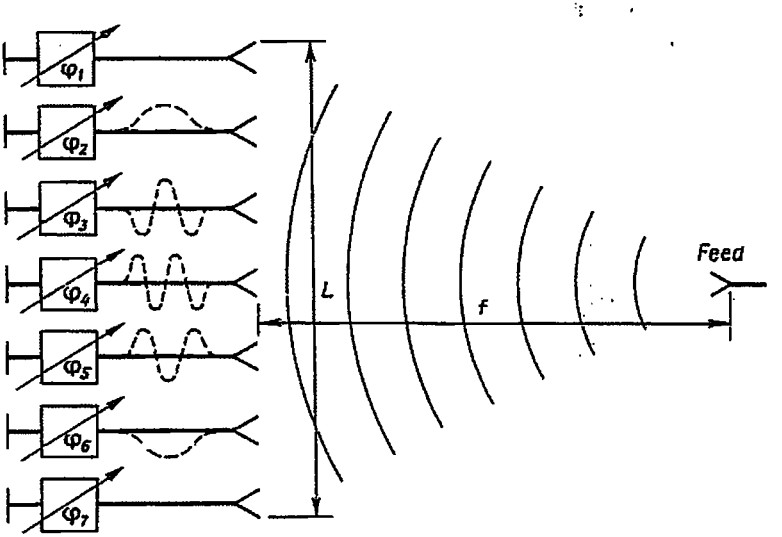


Fig. 15.5. Reflecting optical-type feeder system for the radiators of a phased array

enclosing the radiating system in a large horn extending from the primary feed to the receiving array or by arranging it in the form of a parabolic-cylinder reflector antenna enclosed from all sides by metal walls.

In construction, the reflecting type of phased arrays compares favourably with the feed-through type on several counts. Among other things, any phase shifter is readily accessible from the rear of the array, thus simplifying installation and maintenance. Also, reflecting phase shifters are simpler in design than their feed-through counterparts. On the other hand, the non-reflecting type of phased arrays offers a number of advantages as regards the electrical characteristics, namely: (1) the collecting and radiating arrays can be optimized separately (each may use elements differing in type and placement from those in the other), and (2) the feed does not block the aperture and there is no array reaction to the primary feed.

Phased arrays using unguided (optical) feeder systems can successfully be designed, using the techniques and procedures of prim-

ary-feed optimization developed for lens and reflector antennas. Many factors that have a vital bearing on the aperture efficiency of reflector and lens antennas remain in force in this case as well.

15.3. Phasing Control of Electronically Scanned Arrays

If we apply the phasing equations (see Sec. 12.6) to flat (planar) arrays, we will find that the required phase difference for the radiators located at aperture points with coordinates x_n, y_n may be written as

$$\Phi(x_n, y_n) = -\beta (x_n \sin\theta_0 \cos\varphi_0 + y_n \sin\theta_0 \sin\varphi_0) \tag{15.1}$$

where θ_0, φ_0 is the direction of maximum radiation or the phasing direction.

In order to rapidly determine the phase differences between all the radiators in a rectangular-grid phased array it is convenient first to calculate the phase difference between adjacent elements along the x- and y-axes

$$\Delta\Phi_x = -(2\pi/\lambda)\, d_x \sin\theta_0 \cos\varphi_0$$

and

$$\Delta\Phi_y = -(2\pi/\lambda)\, d_y \sin\theta_0 \sin\varphi_0$$

Following that, these phase differences are consecutively multiplied by all whole numbers m and n up to maximum numbers M and N, corresponding respectively to the number of columns and rows in the array.

Finally, the multiple values thus found are added together to give the phase for each element

$$\Phi_{mn} = m\Delta\Phi_x + n\Delta\Phi_y$$

where m is the mth column and n is the nth row, determining between them the position of each radiator in the array.

The above procedure of phase computation offers an opportunity to build a very elegant array control system (Fig. 15.6) usually known as the *row-and-column control system*. The (mn)th element is energized over two independent channels by two signals embodying the coded values of $m\Delta\Phi_x$ and $n\Delta\Phi_y$. The adders placed next to the phasing devices add together $m\Delta\Phi_x$ and $n\Delta\Phi_y$. The total number of control buses (that is, the total number of control signals) in a phased array with $M \times N$ elements is $M + N$, and this makes the control system simple and reliable and provides for a high speed of response.

If the row-and-column control system in x-y coordinates were directly applied to a phased array in which the elements are located

at the nodes of a triangular grid, this would necessitate the use of twice as many control buses as compared with a rectangular-grid phased array. The number of buses might be reduced by skewing the rows and columns, but this would complicate the computation

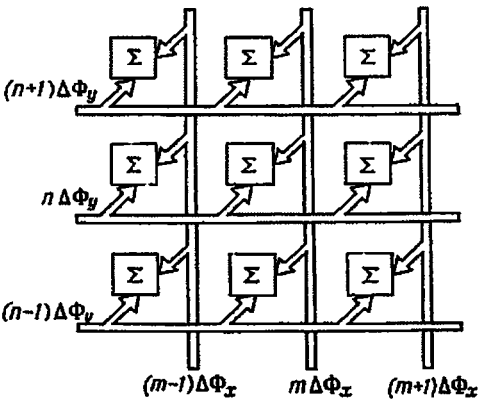


Fig. 15.6. Phasing control by rows and columns

of control codes. In unequally spaced arrays the row-and-column type of control is ineffectual, and the required phase has to be computed for each element individually.

Binary (or step) phasing of array elements. In this case, a phased array uses binary phasing devices that change the phase of the

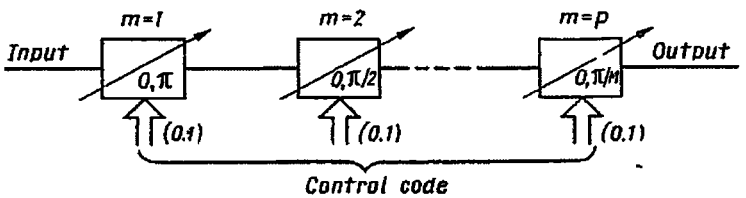


Fig. 15.7. Binary principle of phase control

excitation in each radiator only stepwise by an amount $\Delta\Phi$ called the *phasing step*. As a rule,

$$\Delta\Phi = 2\pi/M$$

where

$$M = 2^p$$

where p is an integer.

As is seen, such phase control systems are based on a binary principle (Fig. 15.7). A feed-through phase shifter is divided into p

stages each of which may reside in any one of two states, the respective phase lag (or lead, as the case may be) being 0 or $\pi/2^{m-1}$, where m refers to the mth stage. In order to select any of $M = 2^p$ likely phase states of the phase shifter, it will suffice to use just p control signals each capable of taking on a value 0 or a value 1.

An important advantage of binary phasing devices as compared with continuous (or analog) devices is an improvement in stability. This is because the control devices operate in the switching mode, using two extreme regions of their characteristics. Another advantage of binary phasing is that the phasing devices can conveniently be controlled by a computer.

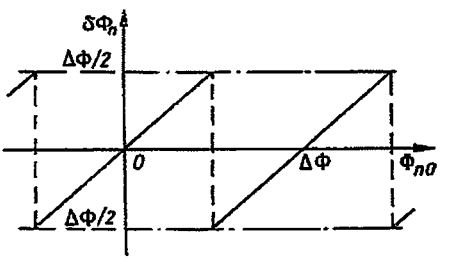


Fig. 15.8. Systematic phase error in an array element with step phase control

When an array is to be phased in a direction θ_0, φ_0, the procedure is first to find for each radiator an exact ('ideal') value of the phase Φ_{n0} by Eq. (15.1) and then to round it off to the nearest discrete value Φ_n

$$\Phi_n = \Delta\Phi E\,(0.5 + \Phi_{n0}/\Delta\Phi) = \Phi_{n0} - \delta\Phi_n$$

where $E\,(x)$ = extraction of the integer part of the number x;
$\delta\Phi_n\,(\Phi_{n0}) = \Phi_{n0} - \Phi_n$ = systematic phase error due to step phasing

The above systematic phase error is a sawtooth function of the ideal phase Φ_{n0} with a period $\Delta\Phi$ (Fig. 15.8).

The array factor of any steps phased array (be it linear, planar or any other arbitrary shape) may, according to Eq. (12.2), be written as

$$f_\Sigma = \sum_{n=1}^{N} |\dot{I}_n| \exp[j(\beta R_n \cos\alpha_n + \Phi_{n0})] \exp(-j\delta\Phi_n),$$

where $R_n \cos\alpha_n$ = path length difference dependent on the coordinates of a radiator and the direction of observation
n = unique sequence of numbers designating the radiators

Let us expand the last exponential function in the above equation into a Fourier series in terms of the variable Φ_{n0}:

$$\exp(-j\delta\Phi_n) = \sum_{m=-\infty}^{\infty} \dot{C}_m \exp[jm(2\pi\Phi_{n0}/\Delta\Phi)] \qquad (15.2)$$

where $\dot{C}_m$'s are the expansion coefficients given by

$$\dot{C}_m = \frac{1}{\Delta\Phi} \int_{-\Delta\Phi/2}^{\Delta\Phi/2} \exp[-jm(2\pi t/\Delta\Phi)] \exp(-jt)\, dt = \frac{(-1)^m \sin(\Delta\Phi/2)}{\Delta\Phi/2 + m\pi} \tag{15.3}$$

Subject to (15.2), the array factor takes the form

$$\dot{f}_\Sigma = \sum_{m=-\infty}^{\infty} \dot{C}_m \dot{F}_{\Sigma m}(\theta, \varphi)$$

$$F_{\Sigma m} = \sum_{n=1}^{N} |\dot{I}_n| \exp\{j[\beta R_n \cos\alpha_n + \Phi_{n0}(1 + 2\pi m/\Delta\Phi)]\} \tag{15.4}$$

On substituting in the above equation the actual values of the path length difference and the ideal phase, it is possible to obtain developed expressions for the array factor of a linear, planar or arbitrary step-phased array. This factor is represented by an infinite sum of partial radiation patterns, $\dot{F}_{\Sigma_m}(\theta, \varphi)$. The largest weight in this sum is assigned to the zero term, $\dot{F}_{\Sigma_0}(\theta, \varphi)$, which is the radiation pattern of an ideally phased array. However, the main maximum of this radiation pattern is reduced by a factor of

$$\dot{C}_0 = [\sin(\Delta\Phi/2)]/(\Delta\Phi/2)$$

which is an indication of a reduction in directivity. The remaining terms in (15.4) give rise to additional unwanted radiation due to step phasing. As the phasing step is progressively decreased, the intensity of the unwanted components tends to zero.

Let us define the direction in which the mth unwanted radiation pattern has a maximum. For this purpose, we should equate the exponent in (15.4) to zero and check if the fields of all the array elements can be combined in phase at the given value of m. For a planar array, the condition

$$\beta R_n \cos\alpha_n + \Phi_{n0}(1 + 2\pi m/\Delta\Phi) = 0$$

is equivalent to

$$\beta \sin\theta_m (x_n \cos\varphi_m + y_n \sin\varphi_m) - \beta \sin\theta_0 (x_n \cos\varphi_0 + y_n \sin\varphi_0)(1 + 2\pi m/\Delta\Phi) = 0 \tag{15.5}$$

where θ_0, φ_0 = phasing direction of the array
x_n, y_n = coordinates of the nth radiator
θ_m, φ_m = direction of maximum radiation of the unwanted partial pattern

Since the condition defined by (15.5) must be satisfied at any x_n, y_n, the solution takes the form

$$\sin \theta_m = (1 + 2\pi m/\Delta\Phi) \sin \theta_0, \quad \sin \theta_0 \geqslant 0$$

$$\varphi_m = \varphi_0 + \delta_m \pi, \quad \delta_m = \begin{cases} 1 & \text{for} \quad m < 0 \\ 0 & \text{for} \quad m \geqslant 0 \end{cases}$$

Thus, the maxima of unwanted partial radiation patterns at any $m \neq 0$ are located in the plane of deflection of the beam from the normal to the aperture, and the unwanted partial radiation patterns themselves are the radiation patterns of a planar array ideally phased in the direction θ_m, φ_m.

According to (15.3), the absolute level of the largest unwanted maxima, for $m = \pm 1$, is $[\sin (\Delta\Phi/2)]/(\pi \pm \Delta\Phi/2)$. For example, when $\Delta\Phi = \pi/2$ (four phase states), the level of the largest unwanted lobe may be 9.6 dB down, and with $\Delta\Phi = \pi/4$, it may be approximately 16.6 dB down. Noticeable grating lobes only occur with linear and planar step-phased arrays. For non-planar phased arrays, Eq. (15.5) does not have a solution that would be the same for all pairs of array elements, and the power of the unwanted radiation patterns is somewhat spread over a wide angular sector. A similar effect can be obtained with linear and planar step-phased arrays as well by creating a nonlinear initial phase taper. Naturally, the initial phase taper should be compensated for by phase shifters in scanning.

The reduction in the aperture efficiency of a step-phased array is given by an approximate formula

$$K_{ae,\,d} = \frac{\sin^2 (\Delta\Phi/2)}{(\Delta\Phi/2)^2}$$

which holds true with and without an initial phase taper.

15.4. Multibeam Arrays

This term applies to antennas with several independent inputs, each corresponding to a radiation pattern, or *beam*, of its own. The simplest example of a multibeam antenna is the Luneberg lens using a cluster of independent feeds (see Fig. 14.11). Monopulse antennas producing sum and difference radiation patterns shaped by a common reflector aperture offer another example of multibeam antennas.

Multibeam antennas can also be built on the basis of arrays using suitable feed-through multiport networks some ports of which are connected to the radiators of the array, whereas the remaining ports correspond to independent channels producing different radiation patterns. The feed-through multiport network used to energize a multibeam array is usually called a *beam-shaping network*. In order to preserve a high antenna efficiency, beam-shaping net-

works are assembled from reactive elements, directional couplers, fixed phase shifters, transmission-line sections, and the like. Sometimes, beam-shaping networks may also include absorbing terminations.

As a rule, a multibeam antenna is designed and built so that it produces a system of sharp beams uniformly filling the specified sector of space. These beams can be switched in any desired manner, thus effecting what we have called amplitude scanning. Beam-position control in amplitude scanning is far simpler than it is with phase scanning because the required amplitudes and phases of excitation for the various beams are 'memorized' in the beam-shaping network, whereas the only job left for the control system is to select the desired radiation pattern.

In systems using multibeam antennas, a possibility exists for signals coming in over various channels to be processed concurrently, and this substantially speeds up the overall operation. Among the demerits of amplitude scanning (beam-switching) are the need to use a relatively elaborate beam-shaping network and a limited capability to generate radiation patterns different from those already available.

Consider the most commonly used beam-shaping networks used to drive equally spaced arrays.

The series beam-shaping network (the Blass matrix). This beam-shaping network consists of two intersecting transmission lines connected at the intersections by loosely coupled directional couplers (Fig. 15.9*a*). The outputs of the vertical lines are connected to the radiators, whereas the inputs of the horizontal and inclined lines are the inputs of the antenna. The free ends of the lines are terminated in matched absorbers. Suppose that the generator is connected to input *1* of the antenna system. R.F. power is propagated on the horizontal line, partly tapped off at each intersection towards the radiators. The radiators are thus excited with a linear phase distribution. The relative phase velocity depends on the length of the delay lines (fixed phase shifters) installed at the inputs of the radiators. If the directional couplers have all the same coupling coefficient, the amplitude of the excitation exponentially tapers off towards the right-hand end of the array, with some power (10% to 20%) being absorbed in the termination. The remaining power goes to shape a sharp beam with its maximum oriented in direction *1* (Fig. 15.9*b*), whereas the other inputs of the antenna system and the remaining matched terminations are left unenergized. Thus, when driven at input *1*, the multibeam antenna in question operates as an ordinary linear array series-fed by the horizontal transmission line via directional couplers.

Now let the generator be connected to input *2*. Since the feeder system starting at input *2* is inclined relative to the first feeder

system, the currents in the array radiators have all a different relative phase velocity because the left-hand radiators are excited with some added phase lag relative to the right-hand radiators. Because of this, the sharp beam produced in space has its maximum oriented in direction *2* (Fig. 15.9*b*).

When the excitation is applied to input *2*, the power propagated in the vertical transmission lines towards the radiators has to pass

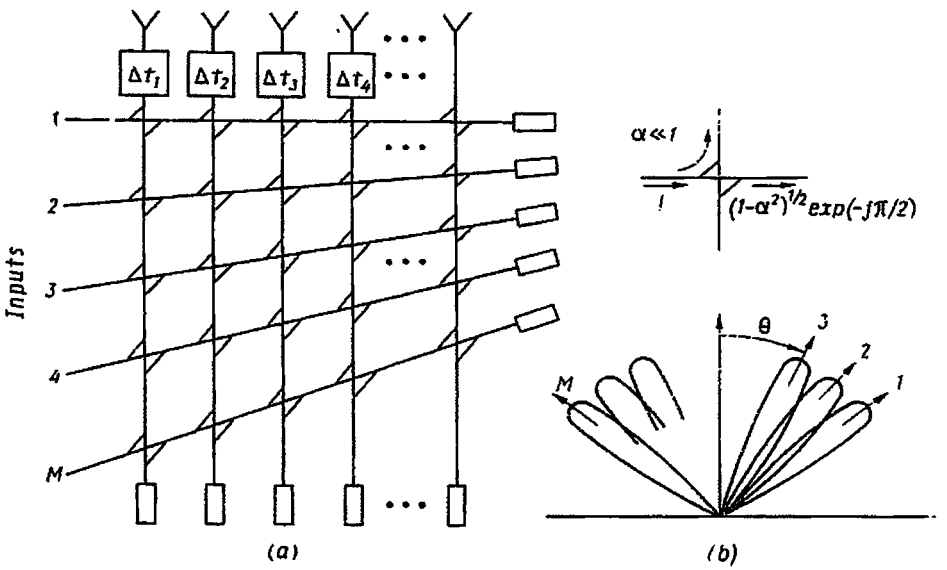


Fig. 15.9. Series-fed multibeam array

through several directional couplers in the upper tier of the beam-shaping network. It would be seen at first glance that the power arriving at the radiators should be attenuated and a wave should appear, travelling in the horizontal feeder system of the beam-shaping network, with some of its power being lost in the load. On this basis, one might expect an impairment in the efficiency of the second beam relative to the first one and, by the same token, an impairment in the third beam relative to the second, and so on. There is, however, a favourable factor which mitigates these undesirable developments. The horizontal feeders in the upper tiers are excited simultaneously by de-phased signals from the lower tiers, the de-phasing being caused by the fact that the horizontal lines are inclined relative to one another. If the peaks of adjacent beams are shifted from one another by a sufficient amount (by at least the half-power beamwidth), the tapped-off de-phased signals will add up in the matched loads of the upper tiers to zero very nearly.

For this reason, the upper tiers do not practically reduce the power conveyed to the radiators from the lower tiers and only produce slight errors in the amplitude and phase distribution of currents in the array radiators. Calculations show that in view of the effect produced by the upper tiers, the amplitude distribution in the array aperture is the same for all inputs, except the top one, whereas the phase error, or the deviation of the phase from a linear distribution, remains negligible, although it increases somewhat for the lower inputs.

The same reasoning could be applied to each of the other inputs.

Thus, the series beam-shaping network offers a means for producing a system of co-existing, but non-interacting sharp beams, provided their number does not exceed the number of array elements (or else there might be a reduction in the efficiency of the system). Series beam-shaping networks are used in ground-based aircraft height finders (with as many as 110 independent beams accommodated within an elevation sector of 0.5 to 40°). Series beam-shaping networks are advantageous in that they can be built with any number of array elements and that the length of the lines to the radiator inputs and between the directional couplers can be chosen such that the beam directions are independent of frequency. On the demerit side, series beam-shaping networks have to use too many directional couplers and a sizeable amount of power is lost in the terminations, thereby reducing the efficiency of the system.

The parallel beam-shaping network (the Butler matrix). What is typical of this arrangement is the fact that r.f. power travels from each input of a multibeam antenna to the array radiators in much the same way as it does in the binary-tier (or herringbone) system. Insight into the operating principle of the parallel beam-shaping network can best be given by an example involving an equally spaced linear array which uses as few as eight radiators (Fig. 15.10*a* and *b*). The basic 'building blocks' of this network are quadrature 3-dB directional couplers. In each coupler, the power coming in from any lower input is split equally among the upper inputs, suffering an additional phase lag of $\pi/2$ as it passes through the 'diagonals' of the directional couplers. In addition to directional couplers, the parallel beam-shaping network uses fixed phase shifters (delay lines), shown as blocks along with the respective phase lags in radians in Fig. 15.10*a*. The electrical length of all the other transmission-line sections may for simplicity be taken equal to zero. By feeding mentally r.f. power to any input of the beam-shaping network and tracing the path travelled by this power to each radiator, it is an easy matter to determine the resultant excitation distribution in amplitude and phase. As an example, Fig. 15.10*a* shows as shaded the directional couplers through which power is conveyed to the radiators when the excitation is applied to the 1st right-

hand (1r) and the 4th left-hand (4l) inputs. In all cases, the input power is split equally among all the array radiators, and all the

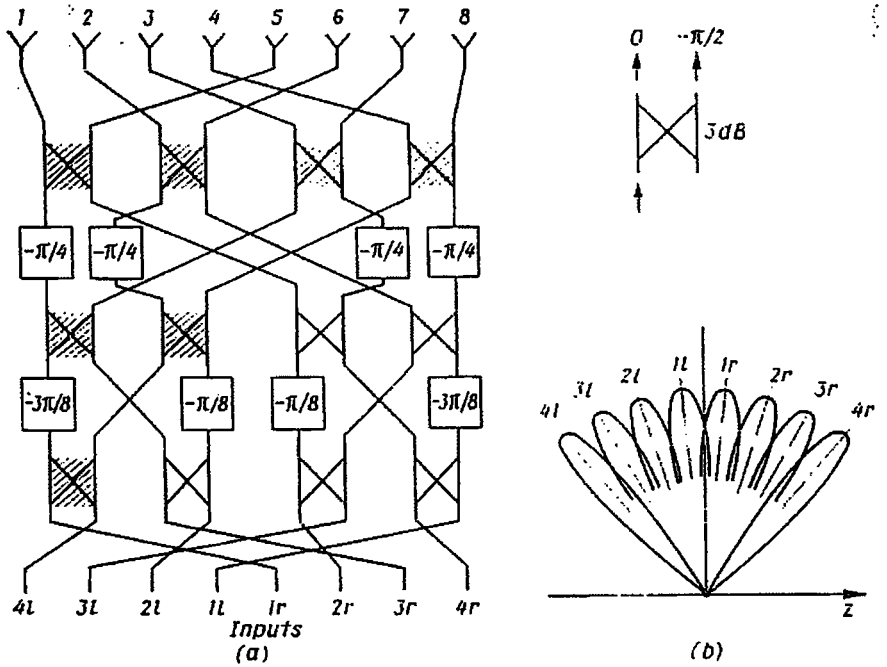


Fig. 15.10. Parallel-fed multibeam array

lower inputs are decoupled from one another and matched (if we neglect the likely reflections from the radiators). In the radiators, the phase tapers off, as listed in Table 15.1 for the four right-hand

Table 15.1

Input number, m	Radiator number, n								
	1	2	3	4	5	6	7	8	$\Delta\Phi_m$
1	$-\frac{5}{8}\pi$	$-\frac{6}{8}\pi$	$-\frac{7}{8}\pi$	$-\frac{8}{8}\pi$	$-\frac{9}{8}\pi$	$-\frac{10}{8}\pi$	$-\frac{11}{8}\pi$	$-\frac{12}{8}\pi$	$\frac{\pi}{8}$
2	$-\frac{5}{8}\pi$	$-\frac{8}{8}\pi$	$-\frac{11}{8}\pi$	$-\frac{14}{8}\pi$	$-\frac{1}{8}\pi$	$-\frac{4}{8}\pi$	$-\frac{7}{8}\pi$	$-\frac{10}{8}\pi$	$\frac{3}{8}\pi$
3	$-\frac{6}{8}\pi$	$-\frac{11}{8}\pi$	$-\frac{16}{8}\pi$	$-\frac{5}{8}\pi$	$-\frac{10}{8}\pi$	$-\frac{15}{8}\pi$	$-\frac{4}{8}\pi$	$-\frac{9}{8}\pi$	$-\frac{5}{8}\pi$
4	$-\frac{8}{8}\pi$	$-\frac{15}{8}\pi$	$-\frac{6}{8}\pi$	$-\frac{13}{8}\pi$	$-\frac{4}{8}\pi$	$-\frac{11}{8}\pi$	$-\frac{2}{8}\pi$	$-\frac{9}{8}\pi$	$\frac{7}{8}\pi$

inputs of the beam-shaping network, excited in turn. The last column in Table 15.1 gives the phase difference $\Delta\Phi_m$ between any two adjacent radiators. The phase advance for the left-hand inputs can readily be deduced on recalling that the right- and left-hand halves of the beam-shaping network are symmetrical.

When the excitation is applied to the mth input of a parallel beam-shaping network, the array factor takes the form

$$F_m = \frac{\sin \Psi_m}{N \sin (\Psi_m/N)}, \quad \Psi_m = 0.5N\beta d\,(\cos\theta - \Delta\Phi_m/\beta d)$$

where N = number of radiators

d = array element spacing

$\beta = 2\pi/\lambda$

θ = angle reckoned from the array axis

The phase difference between two adjacent radiators is, in the general case, given by

$$\Delta\Phi_m = \pm\,[\pi + 2\pi\,(m - 1)]/N, \quad m = 1, 2, \ldots, N/2$$

where the '+' sign is taken for the right-hand inputs, and the '—' sign for the left-hand inputs.

The directions of maximum radiation for the individual beams are found by the following equation:

$$\cos\theta_m = \Delta\Phi_m/\beta d$$

Using the above equations, it can be established that any two beams intersect at a $2/\pi$ level, or 3.92 dB down. Schematically, the positions of the beams produced by an eight-element array are shown in Fig. 15.10*b*.

The parallel beam-shaping network compares favourably with its series counterpart on at least three counts. Firstly, there is no need for absorbing elements. Secondly, the current distribution is ideally equal in amplitude and linear in phase. Thirdly, markedly fewer directional couplers are needed: a mere $(N/2)\cdot\log_2 N$ instead of N^2 in a series beam-shaping network with N beams. A parallel beam-shaping network will, however, be simple only if the number of array elements is $N = 2^p$, where p is an integer. Another disadvantage of parallel beam-shaping networks is the fact that the beams change position in space with changes in frequency.

It is of interest to trace the analogy between the signal-flow graph for the parallel beam-shaping network and the flow diagram for computation by the fast Fourier transform (FFT) algorithms. A study into the properties of the parallel beam-shaping network offers a better insight into the ideas underlying the FFT algorithms. Conversely, the various implementations of these algorithms, including those to a base other than two, suggest some of the approaches that can be taken to building parallel beam-shaping networks.

There is a more general matrix theory of multibeam arrays, outlining the construction of reactive beam-shaping networks for an arbitrary system of interacting radiators. It turns out that in a lossless multibeam antenna with decoupled and matched inputs the radiation patterns of the various channels can only be orthogonal to each other

$$\int\limits_{\Omega=4\pi} \dot{\mathbf{F}}_m \dot{\mathbf{F}}^*_{m'} \, d\Omega = \begin{cases} 0 & \text{for } m \neq m' \\ 4\pi/D_m & \text{for } m = m' \end{cases}$$

where m and m' refer to the mth and the m'th beams, D_m is the directivity of the mth beam, and $d\Omega = \sin\theta d\theta d\varphi$.

15.5. Frequency-Scanned Arrays

In frequency scanning, the phase shift between the excitation currents applied to the radiators and the direction of the array beam are controlled by varying the frequency of the excitation. Most often, frequency-scanned antennas are equally-spaced, series-fed

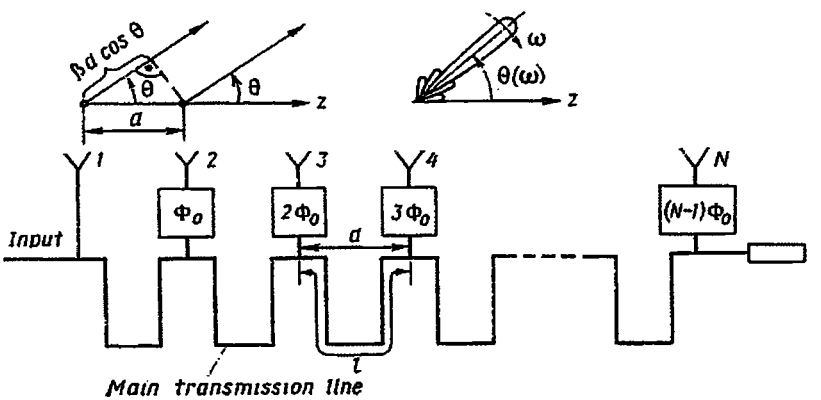


Fig. 15.11. Frequency-scanned linear array

arrays (Fig. 15.11). Power is fed to the radiators in small portions from points on the main feeder line spaced an equal distance l apart. The end of the main feeder line is terminated in a matched load so as to maintain the travelling-wave condition. In Fig. 15.11, the lines going to the radiators contain fixed, frequency-independent phase shifters which take into account the manner of coupling between a given radiator and the feeder line. If the radiators are all the same and coupled to the feeder line in the same manner, there is no need for fixed phase shifters, and they will be omitted from the arrangement. But they will be used, for example, when the

main feeder line is crossed between adjacent elements or when adjacent slots in a rectangular waveguide are excited with a phase reversal.

The array radiators are spaced a distance d apart on the z-axis. In order to avoid multilobe occurrence (see Sec. 11.6), the array element spacing d should be chosen so as to satisfy the inequality

$$d < \lambda_{min}/(1 + |\cos \theta_{1,2}|)$$

where $\theta_{1,2}$ defines the sector of scan, and λ_{min} is the minimum free-space wavelength.

The frequency-dependent phase difference between a pair of adjacent radiators is due to the electrical delay that the excitation wave suffers in the main feeder line over the length l

$$\Delta\Phi(\omega) = \beta(\omega)\, l = \omega l/v_p(\omega)$$

where $\beta(\omega)$ = frequency-dependent phase constant in the main feeder line

$v_p(\omega)$ = phase velocity in the main feeder line

When determining $\beta(\omega)$ in exact terms, one should consider the fact that the radiators load the feeder and that they interact through free space and over the feeder. With a great number of radiators, however, each draws from the feeder only a small fraction of the total propagated power and affects but little the phase velocity. It is then legitimate to deem that the relation $\beta(\omega)$ is the same as it is in a uniform feeder line containing no radiators.

At any frequency, the maximum of array radiation is in the direction $\theta(\omega)$ where the spatial path length difference $\beta_0 d \cos\theta$ between a pair of adjacent elements is equal to the phase lag of the excitation over the length l of the main feeder line plus the additional phase lag Φ_0 due to the coupling devices, and minus the insignificant phase advance which is a multiple of 2π. This condition may be written as the following equality:

$$\beta_0 d \cos[\theta(\omega)] = \beta(\omega)\, l + \Phi_0 - 2\pi M,\ M = 0, \pm 1, \pm 2, \ldots \tag{15.6}$$

As the frequency increases, the phase difference $\beta(\omega)\, l$ rises at a faster rate than does $\beta_0 d$ and this should, in accordance with Eq. (15.6), lead to a rise in $\cos[\theta(\omega)]$, which means that the beam should deviate away from the antenna input.

The rate of change in the spatial position of the main maximum due to changes in frequency is referred to as the *angular frequency sensitivity* of an antenna. To calculate it, we differentiate Eq. (15.6) with respect to frequency. Noting that

(1) $\beta_0 = \omega/c$, where c is the velocity of light in free space,

(2) $\beta = \omega/v_p$, where v_p is the phase velocity of the wave in the main feeder line, and

(3) $\partial\beta(\omega)/\partial\omega = (\partial\omega/\partial\beta)^{-1} = 1/v_g$, where v_g is the group velocity of propagation in the main feeder line, and carrying out simple identity transformations, we obtain the following expression for the angular-frequency sensitivity:

$$\partial\theta \Big/ \frac{\partial\omega}{\omega} = \omega(\partial\theta/\partial\omega) = (0.573/\sin\theta)\left(\cos\theta - \frac{l}{d}\frac{c}{v_g}\right)$$

In the above formula, the factor 0.573 has been introduced for conversion from radians per relative change in frequency to the more convenient degrees per percent of change in frequency. The angular-frequency sensitivity is a maximum when the direction of maximum radiation is along the normal to the array axis

$$\omega \,|\, \partial\theta/\partial\omega|_{max} = 0.573\ (l/d)\ (c/v_g)$$

State-of-the-art microwave generators can readily be retuned within 10% of the operating frequency. In order that with such a frequency agility the beam can be moved within a wide angular sector, the angular-frequency sensitivity must be 5 to 10 deg/% or even greater. The angular-frequency sensitivity can be enhanced in any one of two ways, namely: (1) by using an ordinary feeder line free from a marked frequency dispersion and setting the ratio l/d equal to, say, more than five, or (2) by setting the ratio l/d approximately equal to unity and using in the main feeder line some slow-wave structure with a low relative group velocity, for example $c/v_g > 5$. It can be shown that this takes place at an elevated dispersion (that is, when the phase velocity in the main feeder line strongly depends on frequency) or when the main feeder system is filled with a dielectric of high permittivity. It has been found that any of the two approaches causes an increase in ohmic power losses in the main feeder, whereas the second approach (an increase in the relative group velocity, c/v_g) is additionally accompanied by a reduction in the electric strength of the antenna. Therefore, when choosing the type and dimensions of the main feeder for a frequency-scanned antenna, a compromise decision should be taken that would yield satisfactory results as regards the allowable dimensions of the main feeder and as regards the efficiency of the antenna and the transmitted power.

In designing a frequency-scanned antenna, the first step usually is to select the type of the main feeder and an approximate value for the ratio l/d, proceeding from the required angular-frequency sensitivity, and then to determine the frequency function $\beta(\omega)$ for the feeder type thus chosen. The next step is to select, by resort to Eq. (15.6), such values for d, l, and Φ_0 that at the mid-band frequency of the generator the array beam will be oriented in the desired direction, and finally to refine the frequency dependence of the angular position of the beam.

Direct connection of radiators to the main feeder might entail untoward consequences. Notably, a difficulty might arise in matching the main feeder as the main lobe of the radiation pattern passes through the normal to the array surface, when reflections from all radiators are combined in the main feeder in phase (the 'normal' effect). Also, errors might arise in the amplitude and phase distribution of the excitation due to mutual coupling between the radiators. That is why, as often as not, power is tapped from the main feeder to the radiators with the aid of directional couplers. With a matched load placed in the free arm of each directional coupler, the above drawbacks are avoided due to the absorption of the power reflected from the radiator inputs.

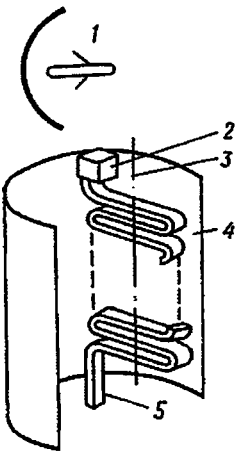


Fig. 15.12. Excitation of a parabolic-cylinder reflector by a frequency-scanned line source: *1*, horn; *2*, load; *3*, focal axis; *4*, reflector; *5*, main transmission system

One practical design of a frequency-scanned antenna is shown in sketch form in Fig. 15.12. Here, the radiation pattern is shaped in the plane of scan by a linear array, whereas in a plane normal to it the beam is shaped due to the optical properties of a parabolic cylinder. The radiators are inclined slots cut in the narrow dimension of a rectangular waveguide which is shaped in a zig-zag fashion as a way of enhancing the angular-frequency sensitivity. The power radiated by the slots is re-directed to the parabolic reflector by a small horn which is shown in the figure in only one view.

Conclusion

It may be said that microwave-component and antenna technology has reached a certain level of maturity. Rather than to look for innovative types of microwave components and antennas, emphasis is placed on improvements in the electrical characteristics, construction and fabrication techniques of existing types and the expansion of their field of application, notably at millimetric and submillimetric wavelengths.

The advance is especially spectacular in the field of multielement antenna systems based on phased arrays. Phased arrays are becoming ever more "intelligent": they are built smart enough to tackle a range of tasks at a time, such as space surveillance, target search, identification and tracking, adaptive suppression of natural or man-made interference, and so on. It is becoming more prevalent to regard the multielement antenna-feeder system as the principal and crucial part of a spatial signal processing system.

A good deal of effort is being put into the development of novel types of phased-array components, such as radiating elements, power dividers, phase shifters, built-in performance monitoring capabilities, etc. There has been a marked progress in the theory and design of low-silhouette phased arrays, that is, those which conform to the surface contour of the vehicles or craft that carry them. Along with general theoretical problems arising in the synthesis of optimal amplitude and phase distributions of phased-array excitation (or, which is the same, the weight coefficients for linear processing of signals of individual radiators), important significance is being attached to the circuits and algorithms that could effect real-time control of phased arrays, notably algorithms providing for adaptation to a varying interference situation.

Another trend in the development of microwave components for phased arrays has been a steady increase in packing density. It is planned to switch over to stripline designs using flexible roll ribbon sandwiched between layers of a honeycomb dielectric. In such designs, the stripline radiators can conveniently be mated to phase

shifters and excitation dividers without resort to intermediate connectors.

Still another rapidly expanding field of activity in antenna technology is the development of hybrid reflector antennas which combine reflectors of large electrical dimensions (in excess of 50 wavelengths) and a radiating array of a small number (not over a hundred) of elements, placed near the focal region of the reflector. This type of antenna is especially promising for use in space communication via geostationary Earth satellites which are required to radiate a system of sharp beams. Importantly, a capability should be provided for each beam in such a system to be pinpointed at the desired receiving station on the ground.

The key theoretical problems related to the development of hybrid reflector antennas include the choice of the best relative position for the reflector and the radiating array and the synthesis of optimum amplitude-phase distributions in the radiating array so as to obtain the desired variable-shape radiation patterns while maximizing the antenna gain. A distinction of on-board hybrid reflector antennas for geostationary communication satellites is the wide use of novel light-weight and very strong structural materials, such as graphite composites, and the fact that an antenna system can be carried to space in a stowed (folded) condition and be unfolded once the carrying spacecraft has been placed in orbit.

Progress in the development of state-of-the-art microwave components and antennas would be unthinkable without resort to computer-aided design, experimental research, and testing. The numerical methods developed in recent years for solving electrodynamic problems offer an opportunity to enhance the accuracy of antenna and microwave component design considerably, thereby cutting down the scope of experimentation that would otherwise be necessary. However, computer-aided design systems for antennas and microwave components are far from completion because the antennas and microwave components currently used widely differ in configuration and, operating frequencies, and also because they are fabricated by as widely differing techniques.

Special emphasis at present is placed on automation in experimental studies. Big strides are being made in novel methods of measurement: from the near field in the antenna aperture or in the transition zone, using suitably designed sources of plane electromagnetic waves, known as collimators; by resort to extra-terrestrial radio sources as in the radio-astronomical method; by use of sources carried on board aircraft as in the fly-over method. Ever more stringent requirements are placed on the quality of measurements. Notably, the dynamic range in determining the sidelobe level of radiation patterns has been extended to 50 or 60 dB. The drive for a reduction in mutual interference has stimulated the

development of special shielded and anechoic enclosures and rooms for measurement purposes.

In the field of measurement automation, a matter of primary concern is the use of computer- and microprocessor-controlled measurement systems operating to predetermined programs. An important role here belongs to the standardization of measurements and the development of suitable primary antenna standards and reference measurement units. Computer-aided quality control of antennas and microwave devices will in the long run cut down their cost by as much as 20 to 25%.

Antenna and microwave component technology is a rapidly expanding division of radio engineering and electronics. Its advances are promoted through the joint effort of experts in applied electrodynamics, systems engineering, radio equipment, electronics, automation, metrology, manufacturing, etc. There is every ground for saying that further advances in antenna and microwave component theory and technology will help greatly to tackle the problems related to global radio communication, space exploration, wireless power transmission over long distances, medical diagnostics, treatment of many diseases, etc.

Appendix
Electrodynamic Foundations of Antenna and Microwave Component Theory

Maxwell's Equations and Field Potentials. Maxwell'c equations, the fundamental equations of electrodynamics, are pivotal to all of the theory of antennas and microwave devices and circuits. If we take a homogeneous and isotropic medium in some regions of which there is a distribution of electric and magnetic currents with a time dependence of the exp ($j\omega t$) form, Maxwell's equations in the abbreviated vector form may be written as

$$\left.\begin{aligned} \operatorname{curl} \dot{\mathbf{H}} &= j\omega \dot{\varepsilon}_a \dot{\mathbf{E}} + \dot{\mathbf{J}}^e \\ \text{and} \\ \operatorname{curl} \dot{\mathbf{E}} &= -j\omega \dot{\mu}_a \dot{\mathbf{H}} - \dot{\mathbf{J}}^m \end{aligned}\right\} \qquad \text{(A.1)}$$

where $\dot{\mathbf{E}}$ = complex amplitude vector of the electric field strength, V/m

$\dot{\mathbf{H}}$ = complex amplitude vector of the magnetic field strength, A/m

$\dot{\varepsilon}_a = \varepsilon_0 \varepsilon \left(1 - j \dfrac{\sigma^e}{\omega \varepsilon_a}\right)$ = complex dielectric constant of the medium (for vacuum, $\varepsilon_0 = 10^{-9}/36\pi$ F/m)

$\dot{\mu}_a = \mu_0 \mu \left(1 - j \dfrac{\sigma^m}{\omega \mu_a}\right)$ = complex permeability of the medium (for vacuum, $\mu_0 = 4\pi \times 10^{-7}$ H/m)

σ^e = volumetric electric conductivity of the medium, Sm/m

σ^m = volumetric magnetic conductivity of the medium, ohms/m

$\dot{\mathbf{J}}^e$ = complex amplitude vector of volumetric impressed electric current density, A/m^2

$\dot{\mathbf{J}}^m$ = complex amplitude vector of volumetric impressed magnetic current density, V/m^2

Magnetic currents are fictitious entities because magnetic charge has not yet been found to exist in nature. However, their inclusion

in the above equations markedly simplifies the calculations of many devices, notably slot antennas.

Equations (A.1) are usually extended to include two more equations

$$\left.\begin{aligned} \operatorname{div} \dot{\mathbf{E}} &= \dot{\rho}^e/\dot{\varepsilon}_a \\ \operatorname{div} \dot{\mathbf{H}} &= \dot{\rho}^m/\dot{\mu}_a \end{aligned}\right\} \tag{A.2}$$

where $\dot{\rho}^e$ = volumetric electric charge density

$\dot{\rho}^m$ = volumetric magnetic charge density

Equations (A.2) are a corollary of Eqs. (A.1) owing to the continuity of electric and magnetic currents defined by the following equations:

$$\left.\begin{aligned} \operatorname{div} \dot{\mathbf{J}}^e + j\omega\dot{\rho}^e &= 0 \\ \operatorname{div} \dot{\mathbf{J}}^m + j\omega\dot{\rho}^m &= 0 \end{aligned}\right\} \tag{A.3}$$

It follows from Eqs. (A.3) that fields can be specified by giving solely currents because the charges can be defined once the current distribution has been specified.

To solve the Maxwell equations (A.1) for a homogeneous medium it is customary to introduce two auxiliary vector fields: the electric-current vector potential $\dot{\mathbf{A}}^e$ and the magnetic-current vector potential $\dot{\mathbf{A}}^m$. In terms of these potentials, the magnetic and electric intensities are defined as follows:

$$\begin{aligned} \dot{\mathbf{H}} &= \operatorname{curl} \dot{\mathbf{A}}^e - j\omega\dot{\varepsilon}_a\dot{\mathbf{A}}^m + (1/j\omega\dot{\mu}_a) \operatorname{grad} \operatorname{div} \dot{\mathbf{A}}^m \\ \dot{\mathbf{E}} &= -\operatorname{curl} \dot{\mathbf{A}}^m - j\omega\dot{\mu}_a\dot{\mathbf{A}}^e + (1/j\omega\dot{\varepsilon}_a) \operatorname{grad} \operatorname{div} \dot{\mathbf{A}}^e \end{aligned} \tag{A.4}$$

Note that some authors use the potentials $\dot{\mathbf{A}}^e$ and $\dot{\mathbf{A}}^m$ which differ from those defined above by additional factors $\dot{\mu}_a$ and $\dot{\varepsilon}_a$. These factors are constants because the medium is deemed homogeneous on introducing the potentials. If one drops the factors $\dot{\mu}_a$ and $\dot{\varepsilon}_a$, the advantages offered by the potentials are not lost, but design relations are simplified.

On substituting (A.4) in (A.1), we obtain the following inhomogeneous vector Helmholtz equations in potentials:

$$\begin{aligned} \nabla^2\dot{\mathbf{A}}^e + \dot{\gamma}^2\dot{\mathbf{A}}^e &= -\dot{\mathbf{J}}^e \\ \nabla^2\dot{\mathbf{A}}^m + \dot{\gamma}^2\dot{\mathbf{A}}^m &= -\dot{\mathbf{J}}^m \end{aligned} \tag{A.5}$$

where

$$\dot{\gamma} = \omega (\dot{\varepsilon}_a \dot{\mu}_a)^{1/2} = \beta - j\alpha$$

is the propagation constant.

Thus, the integration of Maxwell's equations reduces to solving the inhomogeneous vector equations (A.5).

Uniqueness of the Solution and Boundary Conditions. The solutions of Maxwell's equations are unique if they

(1) are finite in all regions that contain no δ-sources;

(2) satisfy the condition of radiation at infinity (the Sommerfeld radiation condition);

(3) satisfy the applicable boundary conditions at the interfaces;

(4) satisfy the Meixner edge condition (which states that as the point of observation tends to sharp edges at the interfaces, some field components should tend to infinity in a certain definite manner).

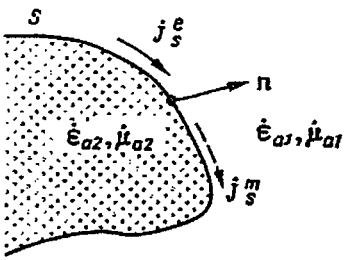


Fig. A.1. Boundary between two media

At the interface between two homogeneous media with a unit normal vector n (Fig. A.1), the boundary conditions reduce to the following equalities:

$$[\mathbf{n}, \dot{\mathbf{E}}_1] = [\mathbf{n}, \dot{\mathbf{E}}_2]$$

$$[\mathbf{n}, \dot{\mathbf{H}}_1] = [\mathbf{n}, \dot{\mathbf{H}}_2]$$

$$\mathbf{n}\dot{\mathbf{D}}_1 = \mathbf{n}\dot{\mathbf{D}}_2$$

$$\mathbf{n}\dot{\mathbf{B}}_1 = \mathbf{n}\dot{\mathbf{B}}_2$$

which express the continuity of the tangential components of the $\dot{\mathbf{E}}$ and $\dot{\mathbf{H}}$ fields and the continuity of the normal components of the induction vectors $\dot{\mathbf{D}} = \varepsilon_a \dot{\mathbf{E}}$ and $\dot{\mathbf{B}} = \dot{\mu}_a \dot{\mathbf{H}}$.

In the presence of *surface currents* and charges at the interface S, the boundary conditions take the form

$$\begin{aligned} [\mathbf{n}, (\dot{\mathbf{H}}_1 - \dot{\mathbf{H}}_2)] &= \dot{\mathbf{J}}_S^e \\ [\mathbf{n}, (\dot{\mathbf{E}}_1 - \dot{\mathbf{E}}_2)] &= -\dot{\mathbf{J}}_S^m \\ \mathbf{n}(\dot{\mathbf{D}}_1 - \dot{\mathbf{D}}_2) &= \dot{\rho}_S^e \\ \mathbf{n}(\dot{\mathbf{B}}_1 - \dot{\mathbf{B}}_2) &= \dot{\rho}_S^m \end{aligned} \qquad \text{(A.6)}$$

It follows from Eqs. (A.6) that at the surface of an ideal conductor

$$[\mathbf{n},\ \dot{\mathbf{E}}_1] = 0$$

$$[\mathbf{n},\ \dot{\mathbf{H}}_1] = \dot{\mathbf{J}}_S^e$$

$$\mathbf{n}\dot{\mathbf{D}}_1 = \dot{\rho}_S^e$$

In the case of a non-ideal electric conductor ($\sigma_2^e \gg \omega\varepsilon_{a2}$), the electromagnetic field penetrates to a depth given by

$$\delta = (2/\omega\mu_{a2}\sigma_2^e)^{1/2}$$

If the radius of curvature of a non-ideal conductor is many times the quantity δ, the following approximate Leontovich boundary condition is valid:

$$[\mathbf{n},\ \dot{\mathbf{E}}_1] = \dot{Z}_w^{\mathrm{Me}}[\mathbf{n},\ [\mathbf{n},\ \dot{\mathbf{H}}_1]]$$

where

$$\dot{Z}_w^{\mathrm{Me}} \approx (j\omega M_{a2}/\sigma_2^e)^{1/2}$$

is the complex wave impedance of the conductor.

The Sommerfeld radiation condition consists basically in that the electromagnetic disturbance due to excitation sources should move to infinity in the form of travelling waves. At the same time, there should be no waves moving from infinity towards the excitation sources.

The Umov-Poynting Theorem*. This theorem is important in calculating the energy characteristics of microwave devices and antennas. In complex notation, this theorem can be written as

$$\frac{1}{2}\int_S [\dot{\mathbf{E}},\ \dot{\mathbf{H}}^*]\,\mathbf{n}\,dS + j\omega\int_V \left\{\frac{\mu_a}{2}\,|\,\dot{\mathbf{H}}\,|^2 - \left(\frac{\varepsilon_a}{2}\right)|\,\dot{\mathbf{E}}\,|^2\right\} dV + \int_V \frac{\sigma^e}{2}\,|\,\mathbf{E}\,|^2\,dV = \frac{1}{2}\int_V (-\dot{\mathbf{J}}_m\dot{\mathbf{H}}^* - \dot{\mathbf{J}}^{e*}\mathbf{E})\,dV \qquad (A.7)$$

where $\mathbf{n}$ is the outward normal to the surface S enclosing the volume V that contains the excitation sources.

The right-hand side of Eq. (A.7) defines the complex power drawn from the generator. The first term on the left-hand side of the equation gives the complex power leaving the volume V, the second term defines the reactive power within the volume V, and the third term defines the power dissipated in the volume V as heat. The density of the power flux leaving the volume V is given by the complex

* Outside the Soviet Union, this is known as the *Poynting theorem.—Translator's note.*

Poynting vector

$$\dot{\Pi} = [\dot{E}, \dot{H}^*]/2$$

Here, as in Eq. (A.7), the factor 1/2 is included for conversion from amplitude (peak) to rms values.

The Permutational Duality Principle. Consider Maxwell's equations in the presence of only the volumetric density of impressed electric currents and in the presence of only the volumetric density of impressed magnetic currents. It is an easy matter to see that these equations (and the equations for the vector potentials and their solutions) can be readily interconverted on applying the following permutations:

$$\dot{E} \rightleftharpoons \dot{H}, \quad \dot{\varepsilon}_a \rightleftharpoons -\dot{\mu}_a, \quad \dot{J}^e \rightleftharpoons -\dot{J}^m, \quad (\dot{\mu}_a/\dot{\varepsilon}_a)^{1/2} \rightleftharpoons -(\dot{\varepsilon}_a/\dot{\mu}_a)^{1/2} \tag{A.8}$$

Resort to the principle of permutational duality makes it unnecessary to solve some electrodynamic problems. For example, if we have found the electromagnetic field due to impressed electric currents having a specified distribution, we need not solve the problem of finding the field due to impressed magnetic currents having the same distribution—it would suffice to apply the principle of permutational duality.

One ought not to forget that the $\dot{E}$ and $\dot{H}$ fields *should simultaneously be interchanged in the boundary conditions as well*, and the wave impedance of the medium should be converted to its wave admittance according to the rule

$$(\dot{\mu}_a/\dot{\varepsilon}_a)^{1/2} \rightleftharpoons -(\dot{\varepsilon}_a/\dot{\mu}_a)^{1/2}$$

which preserves the direction of energy propagation.

The Reciprocity Theorem. This theorem relates the electromagnetic fields $\dot{E}_1$, $\dot{H}_1$ set up by sources in the form of impressed currents $\dot{J}_1^e$, $\dot{J}_1^m$ to the electromagnetic fields $\dot{E}_2$, $\dot{H}_2$ set up by sources in the form of impressed currents $\dot{J}_2^e$, $\dot{J}_2^m$. For a volume V bounded by a surface S and containing at least a part of the two sets of sources, the most common expression for the reciprocity theorem takes the form

$$\int_S [\dot{E}_1, \dot{H}_2] - [\dot{E}_2, \dot{H}_1]\, n\, dS = \int_V (\dot{E}_2\dot{J}_1^e - \dot{E}_1\dot{J}_2^e - \dot{H}_2\dot{J}_1^m + \dot{H}_1\dot{J}_2^m)\, dV \tag{A.9}$$

assuming that the two fields exist in the same isotropic medium. If the surface S encloses all of the sources completely, the integral on the left-hand side of (A.9) vanishes. If the surface S does not

enclose any of the sources, the integral on the right-hand side of (A.9) vanishes.

The reciprocity theorem explains the symmetry of the scattering, impedance and admittance matrices of reciprocal multiport junctions, underlies the reciprocity of transmitting and receiving antennas, and is widely resorted to in setting up integral equations with which current distributions can be found in the radiating systems of antennas.

The theorem may be extended to anisotropic media as well, if one assumes that the permeability and permittivity tensors of the media, on replacing a field $\dot{\mathbf{E}}_1$, $\dot{\mathbf{H}}_1$ by a field $\dot{\mathbf{E}}_2$, $\dot{\mathbf{H}}_2$, are interchanged (transposed) as well (physically this is equivalent to a reversal of

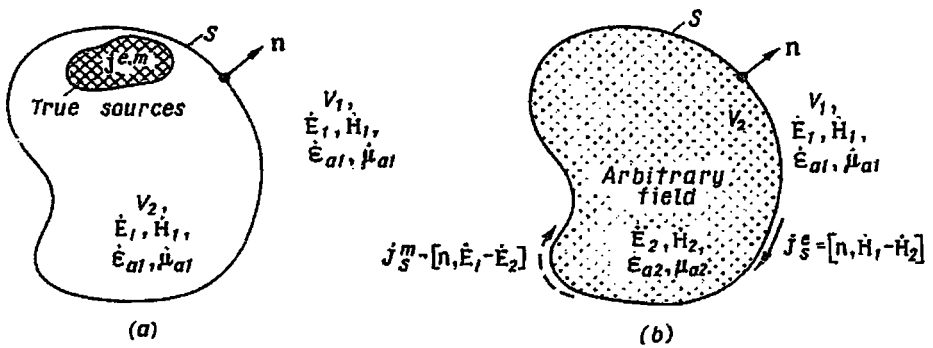


Fig. A.2. Illustration of a field equivalence principle: *a*, original problem; *b*, equivalent electrodynamic problem for region V_1

the biasing field for anisotropic ferrites and plasmas). There is also a more elaborate form of the reciprocity theorem where the field $\dot{\mathbf{E}}_1$, $\dot{\mathbf{H}}_1$ has a frequency ω_1, and a field $\dot{\mathbf{E}}_2$, $\dot{\mathbf{H}}_2$ has a frequency ω_2, such that $\omega_1 \neq \omega_2$.

Field Equivalence Principles. A field equivalence principle refers to the replacement of the actual sources of an electromagnetic field by a set of more convenient equivalent surface currents.

To establish the basic concepts involved, let S be a closed surface which separates space into regions V_1 and V_2 and encloses all the actual sources which radiate an electromagnetic field $\dot{\mathbf{E}}_1$, $\dot{\mathbf{H}}_1$ (Fig. A.2*a*). Let us remove the sources from V_2 and postulate the existence in the region V_2 of an arbitrary field $\dot{\mathbf{E}}_2$, $\dot{\mathbf{H}}_2$ (whose sources satisfying Maxwell's equations for an arbitrary medium $\dot{\varepsilon}_{a2}$, $\dot{\mu}_{a2}$ are located outside V_2). At the same time let us preserve the original field in V_1 (Fig. A.2*b*). In order that the total field throughout V_1 and V_2 be a valid solution of Maxwell's equations it is necessary

to postulate also the existence of equivalent sources on the surface S in order to properly join the two solutions together. The equivalent sources on S must be so chosen as to satisfy the boundary conditions defined in (A.6). Thus we require electric and magnetic surface sources

$$\dot{\mathbf{J}}_S^e = [\mathbf{n}, (\dot{\mathbf{H}}_1 - \dot{\mathbf{H}}_2)]$$

$$\dot{\mathbf{J}}_S^m = -[\mathbf{n}, (\dot{\mathbf{E}}_1 - \dot{\mathbf{E}}_2)] \tag{A.10}$$

The fictitious surface currents $\dot{\mathbf{J}}_S^e$ and $\dot{\mathbf{J}}_S^m$ may be treated as the equivalent sources radiating the $\dot{\mathbf{E}}_1$, $\dot{\mathbf{H}}_1$ field in V_1 and the $\dot{\mathbf{E}}_2$, $\dot{\mathbf{H}}_2$ field in V_2. This is the essence of the equivalence principle. Note that the original field sources are removed from V_2. By the same token, the sources radiating the $\dot{\mathbf{E}}_2$, $\dot{\mathbf{H}}_2$ field are removed from V_1. Importantly, the new $\dot{\mathbf{E}}_2$, $\dot{\mathbf{H}}_2$ field and the medium $\dot{\varepsilon}_{a2}$, $\dot{\mu}_{a2}$ may be formed at the investigator's discretion independently of the $\dot{\mathbf{E}}_1$, $\dot{\mathbf{H}}_1$ field and the medium $\dot{\varepsilon}_{a1}$, $\dot{\mu}_{a1}$.

Most commonly, the equivalent sources are chosen as follows.

1. The $\dot{\mathbf{E}}_2$, $\dot{\mathbf{H}}_2$ field is postulated to be zero, but the medium is preserved unchanged, that is, such that $\dot{\varepsilon}_{a2} \equiv \dot{\varepsilon}_{a1}$ and $\dot{\mu}_{a2} \equiv \dot{\mu}_{a1}$. Then Eq. (A.10) takes the form

$$\dot{\mathbf{J}}_S^e = [\mathbf{n}, \dot{\mathbf{H}}_1]$$

$$\dot{\mathbf{J}}_S^m = -[\mathbf{n}, \dot{\mathbf{E}}_1] \tag{A.11}$$

and the electromagnetic field in V_1 may be regarded as being due to the electric and magnetic equivalent currents defined by Eq. (A.11), that exist in an unbounded homogeneous medium.

2. The $\dot{\mathbf{E}}_2$, $\dot{\mathbf{H}}_2$ field is postulated to be zero, and the medium within V_2 is assumed to possess the properties of an ideal conductor. In the circumstances, the equivalent surface electric current $\dot{\mathbf{J}}_S^e$ is 'non-radiating' (it is 'shorted', as it were, by the medium inside V_2), and the field in V_1 may be regarded as being due solely to the equivalent magnetic currents $\dot{\mathbf{J}}_S^m = -[\mathbf{n}, \dot{\mathbf{E}}_1]$, but on the condition that the presence of the ideally conducting body V_2 is taken into account.

3. The $\dot{\mathbf{E}}_2$, $\dot{\mathbf{H}}_2$ field is postulated to be zero, and the medium in V_2 is assumed to be an ideal magnetic. In such a case, it is the equivalent surface magnetic current that is a non-radiating one, and the field in V_1 may be regarded as being due solely to the equivalent

electric currents $\dot{J}_S^e = [\mathbf{n}, \dot{\mathbf{H}}_1]$, but on the condition that the presence of an ideally magnetic body, V_2, subject to the boundary condition $\dot{\mathbf{H}}_t = 0$, is taken into account.

Methods (2) and (3) are especially convenient in calculating the radiation from apertures in infinite flat flanges which are treated using the image method (see below). Apart from serving as a theoretical basis for the design of antennas and microwave devices, the field equivalence principle is directly used in various holographic and radioholographic devices. Notably, the field equivalence principle lies at the basis of a rapidly developing method used to synthesize the radiation patterns of antennas from the measured distributions of the tangential components of the near field on a suitable (planar, cylindrical or spherical) surface.

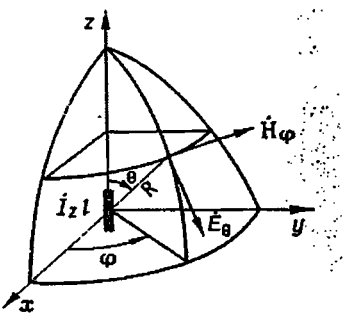


Fig. A.3. Elementary dipole

The success in the use of the field equivalence principle in theoretical applications is due to the fact that it offers an opportunity to determine the distributions of equivalent surface currents on suitable surfaces enclosing real radiators. As a rule, an exact calculation of such distributions is too time- and effort-consuming, and this forces one to resort to various approximate distributions of equivalent sources, ordinarily on the basis of concepts from geometrical or physical optics, and also on the basis of physical assumptions about the behaviour of the near field when the construction of the radiator is known.

Radiation from Elementary Dipoles. The elementary sources of an electromagnetic field are dipoles—short elements (substantially shorter than a wavelength) of uniform electric currents $\dot{I}^e$ or magnetic currents $\dot{I}^m$. By virtue of the linearity of Maxwell's equations, the electromagnetic field of any radiating system may always be pictured as a superposition of the fields due to the elementary dipoles.

For a dipole located at the origin of a spherical coordinate system, with the current oriented along the z-axis (Fig. A.3), the exact expressions for the electromagnetic field of dipoles in unbounded free space take the form

$$\dot{\mathbf{H}} = (j\dot{I}^e l/2\lambda)\, \mathbf{i}_\varphi \sin\theta\, (1 - j/\beta R)\, [\exp(-j\beta R)/R]$$

$$\dot{\mathbf{E}} = (jZ_m \dot{I}^e l/2\lambda)\, \{\mathbf{i}_\theta \sin\theta\, [1 - j/\beta R - 1/(\beta R)^2] - \mathbf{i}_R \cos\theta\, [j/\beta R + 1/(\beta R)^2]\}\, [\exp(-j\beta R)/R] \quad \text{(A.12a)}$$

for an *electric dipole*, and

$$\dot{\mathbf{E}} = (-j\dot{I}^m l/2\lambda)\, \mathbf{i}_\varphi \sin\theta\, (1 - j/\beta R)\, [\exp(-j\beta R)/R]$$

$$\dot{\mathbf{H}} = \{(j\dot{I}^m l/2Z_m\lambda)\, \mathbf{i}_\theta \sin\theta\, [1 - j/\beta R - 1/(\beta R)^2] - \mathbf{i}_R \cos\theta\, [j/\beta R + 1/(\beta R)^2]\}\, [\exp(-j\beta R)/R] \quad \text{(A.12b)}$$

for a *magnetic dipole.*

Here,

$\mathbf{i}_\theta$, $\mathbf{i}_\varphi$ = unit vectors in the spherical coordinate system (R, θ, φ)

$$\beta = \omega(\mu_0\varepsilon_0)^{1/2} = 2\pi/\lambda$$

$$Z_m = (\mu_0/\varepsilon_0)^{1/2} = 120\pi \text{ ohms}$$

In the far-zone region where $\beta R \gg 1$, Eqs. (A.12) retain only those terms which vary inversely as the first power of the distance R. The far zone has no radial components, whereas the θ and φ components are in phase and give rise to the only real radial component of the Poynting vector. The radial component of the Poynting vector varies as $\sin^2\theta$, which is the power radiation pattern of a dipole source, and θ is the observation angle.

By integrating the Poynting vector over the surface of an arbitrary sphere in the far-zone region, it is possible to compute the radiated power P_Σ. This is known as the *Poynting vector method.* In antenna theory it is customary to express the radiated power in terms of radiation resistance R_Σ and radiation conductance G_Σ. For elementary dipoles the radiation resistance (in ohms) and the radiation conductance (in siemens) are introduced by the following relations:

$$P_\Sigma^e = 0.5\, |\dot{I}^e|^2 R_\Sigma$$

$$P_\Sigma^m = 0.5\, |\dot{I}^m|^2 G_\Sigma$$

and are respectively given by

$$R_\Sigma^e = Z_m (2\pi/3)(l/\lambda)^2 = 80\pi^2 (l/\lambda)^2 \quad \text{(A.13a)}$$

for an *electric dipole*, and

$$G_\Sigma^m = (2\pi/3Z_m)(l/\lambda)^2 = R_\Sigma^e/Z_m^2 \quad \text{(A.13b)}$$

for a *magnetic dipole.*

Knowledge of the radiation resistance (or conductance) of an antenna is important when comparing with the ohmic loss resistance R_{loss} which determines the power P_{loss}, dissipated as the antenna is heated by the radiating currents. In the case of an electric dipole, for example, the ohmic power loss can be estimated by the following equation:

$$P_{loss} = 0.5|\dot{I}^e|^2 R_{loss}$$

where the loss resistance $R_{loss} = R_1 l$ is expressed in terms of the resistance per unit length of the dipole conductor, as found at r.f. by the theory of skin-effect. If we define the efficiency of a dipole as

$$\eta = \frac{P_\Sigma}{P_\Sigma + P_{loss}} = \frac{R_\Sigma}{R_\Sigma + R_{loss}} = \frac{l/\lambda}{l/\lambda + 3R_1\lambda/2\pi Z_m} \tag{A.14}$$

it will be seen that, with the ohmic resistance per wavelength fixed in advance, the efficiency can only be enhanced by using a longer dipole. *If, on the other hand, the length of the dipole made of a real conductor is progressively decreased, the efficiency will tend to zero.* This tendency for the efficiency to go down as the electrical dimension is reduced is inherent in all real radiators of small electrical dimensions.

Radiation from a loop antenna. As has been noted, magnetic currents are fictitious quantities. It has been found, however, that from the view-point of the behaviour of the electromagnetic field at some distance, a loop antenna of a size far smaller than the wavelength will behave similarly to a magnetic dipole. The equivalent magnetic moment $\dot{I}^m l$ of a small, single-turn loop antenna of an arbitrary shape and of area S, threaded by a constant-amplitude r.f. current $\dot{I}_l^e$ is given by

$$\dot{I}^m l = j\dot{I}^e Z_m \beta S \tag{A.15}$$

The above relation offers an opportunity to determine the electromagnetic field at some distance from a loop antenna with the aid of Eqs. (A.12b). Using Eqs. (A.13) and (A.15), it is possible to determine the radiation resistance (in ohms) for a single-turn loop antenna

$$R_\Sigma^l = 8\pi^3 Z_m S^2/3\lambda^4 = 320\pi^4 S^2/\lambda^4$$

If a loop antenna has several turns, its radiation resistance per current in one turn, that is, per input current, increases n^2 times.

By analogy with a loop carrying an electric current, we may introduce the concept of a loop carrying a magnetic current. Invoking the principle of permutational duality defined by Eq. (A.8), it can be established that such a loop is equivalent to an electric dipole for which the electric moment is

$$\dot{I}^e l = -j\dot{I}_l^m \beta S/Z_m$$

where $\dot{I}_l^m$ = uniform magnetic current traversing the loop
S = loop area
$Z_m = (\mu_0/\varepsilon_0)^{1/2}$ = wave impedance of the medium

The Image Method. Let elementary field sources be located at a height h above an ideally conducting, infinite, thin metal sheet

(Fig. A.4). The electromagnetic field radiated by the sources induces surface electric currents in the sheet. In the upper and lower half-spaces these currents give rise to secondary fields which are symmetrical about the sheet. The ideally conducting sheet screens the lower half-space, and the induced currents are distributed in such a way that the secondary field at any point in the lower half-space is equal in amplitude and opposite in phase to the primary field set up by the sources in that half-space.

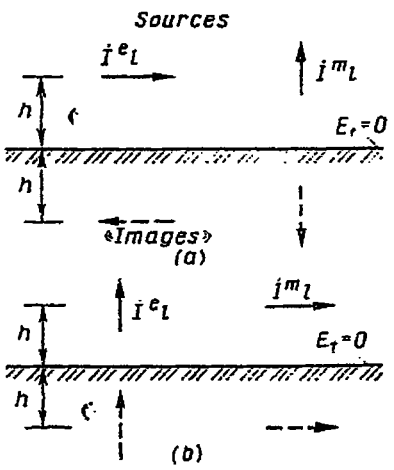


Fig. A.4. Directions of currents in images

By the image method the secondary field in the upper half-space will not change if we remove the conducting sheet and place at the image point a source carrying a current equal in amplitude to the current in the actual source and having a direction such that the tangential component of the total electric field at the surface of the sheet is zero. For a horizontal electric dipole and a vertical magnetic dipole the current in the image must flow in a direction opposite to that of the current in the actual source (Fig. A.4*a*). For a vertical electric dipole and a horizontal magnetic dipole the current in the image must flow in the same direction as it does in the actual source (Fig. A.4*b*).

If the height h above the sheet is zero, the primary and secondary fields radiated by a horizontal electric dipole and a vertical magnetic dipole will be equal in magnitude and opposite in phase, the total field becomes equal to zero, and there is no radiation. Conversely, with a vertical electric dipole and a horizontal magnetic dipole, the primary and secondary fields at $h = 0$ become equal in magnitude and sign, so that the total field is twice as great as the field due to the same source in free space. As to the radiation resistance and conductance of the sources at $h = 0$, they become equal to zero in the former case and are doubled in the latter. The doubling of the radiation resistance and conductance is due to the fact that the density of the radiated power at each point in space is quadrupled, but the power is radiated only into the upper half-space.

The Elementary Slot Antenna. A horizontal magnetic dipole arranged at the surface of an ideally conducting plane behaves practically as a radiating slot cut in a conducting sheet. The magnetic current of a horizontal dipole is nothing but the potential difference between the edges of the slot. The ratio of the power radiated by the slot to the square of the rms voltage across the slot

gives the radiation conductance of an elementary slot antenna. Obviously, this conductance can be found by Eq. (A.13b) multiplied by two.

The Huygens Source. This radiator is an equivalent source which replaces an element of the front of a plane electromagnetic wave propagated in free space. Suppose that the wave is propagated along the z-axis and the wave front element $dxdy$ is located at the origin

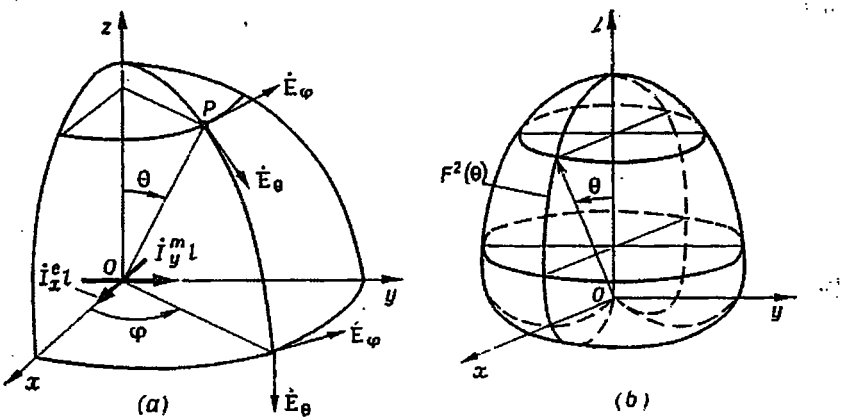


Fig. A.5. Huygens source (a) and its radiation pattern (b)

of a spherical coordinate system. According to Eq. (A.11), the element of the wave front of a plane wave may be replaced by perpendicular electric and magnetic dipoles for which the equivalent moments are

$$\dot{I}^e_x l = -\dot{H}_y dxdy$$

and

$$\dot{I}^m_y l = -\dot{E}_x dxdy$$

where $\dot{E}_x$ and $\dot{H}_y = \dot{E}_x/Z_m$ are the field intensities of a linearly polarized plane wave at the origin of coordinates, that is, at the position of the Huygens source (Fig. A.5). Using Eq. (A.12) for the fields of elementary dipoles and carrying out the necessary coordinate conversion related to the position of the dipole axes, we obtain the following expressions for the *far field* ($R \gg dx$, dy) of the Huygens source:

$$d\dot{\mathbf{E}} = (\mathbf{i}_\theta \cos\varphi - \mathbf{i}_\varphi \sin\varphi)\,(j\dot{E}_x/2\lambda)\,(1 + \cos\theta) \times [\exp(-j\beta R)/R] \qquad (A.16)$$

where $\mathbf{i}_\theta$, $\mathbf{i}_\varphi$ are unit vectors in the spherical coordinate system,

The radiation from the Huygens source is linearly polarized because the components $d\dot{E}_\theta$ and $d\dot{E}_\varphi$ are in phase.

The components of the far magnetic field are given by the obvious relations

$$d\dot{H}_\varphi = d\dot{E}_\theta / \dot{Z}_m$$

and

$$d\dot{H}_\theta = -d\dot{E}_\varphi / Z_m$$

where

$$Z_m = (\mu_0/\varepsilon_0)^{1/2}$$

The spatial power radiation pattern of the Huygens source does not depend on the angle φ (by virtue of the identity $\cos^2 \varphi + \sin^2 \varphi = 1$), and in any plane passing through the z-axis it is the square of a cardioid pattern

$$F^2(\theta) = (1 + \cos \theta)^2/4$$

Note that the radiation is a maximum in the positive z-direction and there is no radiation in the negative z-direction. Thus, the Huygens source is an elementary unidirectional radiator.

Bibliography

1. Markov G.T., Sazonov D.M. *Antennas.* A Textbook for Colleges. 2nd Edition. Moscow, Energiya, 1975 (in Russian).
2. Sazonov D.M., Gridin A.N., Mishustin B.A. *Microwave Circuits.* Mir Publishers, Moscow, 1982 (in English).
3. Voskresensky D.I., ed. *Antennas and Microwave Devices. Design of Phased Arrays.* Moscow, Radio i svyaz, 1981 (in Russian).
4. Collin R.E., Zucker F. *Antenna Theory.* McGraw Hill. New York, 1969.
5. Zeitlin N.M. (ed.). *Measurement of Microwave Antenna Characteristics.* Moscow, Radio i svyaz, 1985 (in Russian).
6. Eisenberg G.Z., ed. *Short-wave Antennas.* Moscow, Radio i svyaz, 1985 (in Russian).
7. Nikolsky V.V. *Computer-Aided Design of Microwave Circuits.* Moscow, Radio i svyaz, 1982 (in Russian).
8. Rudge A.W., Milne K., Olver A.D., Knight P. (editors). *The Handbook of Antenna Design,* Peter Peregrinus, Ltd, 1986.
9. Wohlman V.I., ed. *The Handbook of Microwave Stripline Device Design.* Moscow, Radio i svyaz, 1982 (in Russian).
10. Hizha G.S., Vendik I.B., Serebryakova E.A. *Microwave Phase Shifters and Switches Based on Integrated-Circuit PIN Diodes.* Moscow, Radio i svyaz, 1984 (in Russian).
11. Hansen R.C., ed. *Microwave Scanning Antennas.* Academic Press New York and London. Vol. I, Apertures, 1964; vol. II, Array Theory and Practice, 1966; vol. III, Array Systems, 1967.
12 Helszajn J. *Passive and Active Microwave Circuits.* J. Wiley & Sons, 1978.
13. Monzingo R.A., Miller T.W. *Introduction to Adaptive Arrays.* J. Wiley & Sons, 1980.
14. Mailloux R.J. *Phased Array Theory and Technology.* Proc. of IEEE, vol. 70, No. 3, March, 1982.
15. Strahov A.F. *Automated Antenna Measurements.* Moscow, Radio i svyaz, 1985 (in Russian).
16. Gupta K.C., Ramesh Garg, Rakesh Chadha. *Computer-Aided Design of Microwave Circuits.* Artech House, Inc., 1981.
17. Belyaev B.G. *Integral Field Characteristics of a Focused, Circular Aperture Antenna in the Fresnel Region.* Collection "Antennas", No. 23, Moscow, Svyaz, 1976 (in Russian).

Index

A

Amplitude taper, 332
Antenna arrays, 216
Antenna fields,
  far-zone, 220, 222
  Frauenhofer, 220
  near-zone
  radiation, 220
  transition-zone, 224
Antennas
  annular-slot, 302
  aperture-type, 216, 422
  Archimedes spiral, 312
    radiation pattern of, 314
  asymmetric reflector, 453
  bandwidth of, 241
  basic structure of, 217
  Cassegrain(ian) two-reflector, 450
  classification of, 215
  corrugated-surface, 360
  dielectric-rod, 354
  dipole, 263
  directivity of, 234
  discone, 295
  disc-on-rod, 363
  effective length of, 238
  efficiency of, 239
  elementary slot, 494
  E-plane sectoral horn, 426
  folded-dipole, 291
  frequency-independent, 307
    spiral, 310
  Gregorian two-reflector, 451
  helical, 356
    axial-beam mode, 359
  horn, 423
  horn parabolic reflector, 455
  H-plane sectoral horn, 424
  hybrid reflector, 453
  lens, 427
    constrained-type, 431
    delay type, 429
    dielectric, 429
    Luneberg, 434
    metal-plate, 429
    phase-advance type, 429
    zoning of, 431
  linear, 319
    radiation modes of, 324
  log-periodic, 316
  maximum operating power of, 241
  microstrip, 303
  multibeam, 434
  multiplate, 455
  mutual impedance of, 257
  parabolic reflector, 435
    feeds for, 444
  periscope two-reflector, 455
  phase centre of, 232
  phase characteristic of, 232
  polyrod, 354
  power transfer between, 259
  purpose of, 215
  pyramidal horn, 426
  quasi-paraboloidal two-reflector, 452

radiating length of, 238
receiving, 242
effective area of, 250
load power of, 250
noise temperature of, 251
sensitivity of 252
shaped-beam, 458
slot, 263, 298
in infinite plane screen, 278
spherical reflector, 456
spiral, 310
equiangular, 311
strip, 303
supergain, 420
travelling-wave, 215, 354
turnstile, 296
two-reflector parabolic, 218
Uda-Yagi, 290, 362
Aperture,
focused, 394, 397
planar, 378
analysis of, 380
beam scanning in, 384
in-phase, 372
radiating, 369
Arrays, 319, 369, 459
broadside, 325
electronically scanned, 467
end-fire, 325
frequency-scanned, 477
gain of, 352
input power for, 352
linear, 347
directivity of, 349
synthesis of, 415
multibeam, 471
phased, 387, 461
scanning methods in, 459
slotted waveguide, 363
resonant, 367
slot types in, 363
with rectangular grid, 388
with triangular grid, 389
Yagi-Uda, 362
Attenuation constant, 20
Attenuators, 184
below-cutoff waveguide, 184
strip, 185
variable, 184

B

Beam broadening, 333
Beam position control, 448
Beam-shaping network, 471
Blass matrix, 473
Butler matrix, 474
Beam-switching, 459
Beamwidth, 235
half-power, 235
ideal linear antenna, 325
zero-power, 235
Broadband matching circuits, 151

C

Critical frequency, 19
Cross polarization, 230
Cutoff frequency, 19

D

Dipoles, 263
active, 282
batwing, 296
coupled, 282
directivity of, 269
electric, 263
passive, 282
radiation pattern of, 269
radiation resistance of, 269
Directional couplers
directivity of, 129
reverse-coupling, 131
ring-type, 135
single-hole sidewall, 127
square, 136
Directivity, 234
dipole, 269
ideal linear radiator, 328
linear arrays, 349
Dispersive medium, 19
Duplexers, 185
ATR-tube, 185
dual TR tube balanced, 188
TR-tube, 185

E
Element factor, 286
Equivalent long line, 37
Excitation
  anti-phase, 311, 313
  even, 125
  in-phase, 311, 314
  odd, 125
  with amplitude taper, 332
F
Faraday-effect devices, 202
Faraday-rotation isolators, 203
Feeds, 444
  Cutler, 446
  dipole, 445
  dual-aperture, 446
  line-source, 448
  monopulse, 449
  waveguide horn, 446
Ferrites, 198
Ferrite circulators, 201
  phase-shift, 207
Ferrite devices,
  transverse-biased, 204
Ferrite filters,
  loop-coupled, 213
  magnetically tunable, 212
  microstrip, 213
  rectangular-guide, 213
Ferrite gyrators, 201
Ferrite isolators, 201
  field-displacement, 207
  resonance, 205
  stripline, 207
Field equivalence principle, 489
Field potentials, 484
Fresnel approximation, 394
G
Glating lobes, 344, 387
  suppression of, 343
Gyromagnetic ratio, 199
H
Hallen's equation, 265
Hansen-Woodyard condition for increased directivity, 327
Horn, 423
  conical, 427
  E-plane sectoral, 426
  H-plane sectoral, 423
  pyramidal, 426
Huygens source, 495
Hybrid ring, 136
I
Ideal linear radiator, 319, 322
  directivity of, 328
Image method, 493
Impedance matching transformers, 170
  binomial, 174
  Butterworth, 171
  Chebyshev, 171
  exponential, 176
  stepped, 170
  tapered, 170
Impedance transformation, 48
  by microwave switching diodes, 191
Impedance transformers,
  half-wave, 50
  quarter-wave, 49
L
Larmor frequency, 199
Linear antenna systems,
  synthesis of, 401
Linear arrays,
  synthesis of, 415
Linear radiators,
  synthesis by the Fourier integral method, 407
  synthesis by the partial radiation pattern method, 412
Load immitance measurement, 51
M
Magic-tee, 114
Magnetic dipole, 278
Matched loads, 58
Matching,
  broadband, 55
    by stepped-impedance transformers, 170

by tapered impedance transformers, 175
complex loads, 177
by lumped reactances, 56
by quarter-wave transformer, 55
narrowband, 54
Matching elements,
shunt reactive, 58
Matrices,
admittance, 91
cascading, 118
impedance, 91
scattering, 85, 142
transformation, 118
Maximum transmitted power, 22
Maxwell's equations, 484
Mechanical switching devices, 182
Microwave adapters, 68
series-stub coaxial-to-waveguide, 69
with transverse post, 69
Microwave connectors, 64
Microwave control devices, 181
Microwave couplings, 64
choke flange, 65
cover flange, 65
plane-faced, 65
Microwave ferrite devices, 198
Microwave filters, 150
bandpass, 150
band-rejection, 150
band-stop, 150
Butterworth, 153
Chebyshev, 153
dielectric-resonator, 168
direct-coupled, 167
equal-ripple, 153
high-pass, 150
ladder prototype, 155
low-pass, 150
maximally flat, 153
prototypes, 152
using direct-coupled resonators, 165
using quarter-wave resonators, 165
Microwave multiport junctions, 82
cascading of, 142
composite, 116
matrices of, 83
ideal and real, 98
non-dissipative, 101
reciprocal, 99
symmetrical, 110
Microwave switches, 193
diode-type, 193
Microwave switching diodes, 189
Microwave transducers,
coaxial-to-trip line, 73
waveguide-to-strip line, 74
Microwave transitions, 68
coaxial-to-waveguide, 69
doorknob-type coax-to-waveguide, 70
probe-type, 69
Mode launchers, 72
Mode transducers, 68
Modes,
dominant, 19
principal, 19
TE, 19
TEM, 19
TM, 19
Multilobe ocurrence, 387

N

Noise sources, 255
Non-reciprocal devices, 200
Normalized admittance, 41
Normalized impedance, 41
Nyquist's formula, 252

P

Path length difference, 221
Pattern multiplication theorem, 287
Permutational duality principle, 488
Phase centre, 232
Phase errors, 335
cubic, 337
linear, 335
quadratic, 337
random, 339
Phase shift
differential, 130
non-reciprocal, 207
quadrature, 132

Phase shifters,
  feed-through, coupler-type, 196
    loaded-line, 196
    diode, 195
    mechanical, 182
    multi-state, 197
  ferrite, 210
    Reggia-Spencer, 210
    transverse-biased, 211
  reflecting, 194
  switching-diode, 194
  waveguide, 183
Physical optics approximation, 299
Pocklington's equation, 265
Power attenuation factor, 46
Power dissipation, 21
Power dividers, 139, 461
Product pattern theorem, 287

Q

Q-factor,
  external, 161
  loaded, 161
  unloaded, 161
Quarter-wave skirt, 293

R

Radiating apertures, 369
Radiation
  from a loop antenna, 493
  from elementary dipoles, 491
Radiation centre, 233
Radiation modes, 325
  broadside, 325
  end-fire, 325
Radiation pattern, 226
  cosecant, 227
  difference, 410
  fan-beam, 227
  field-strength, 226
  pencil-beam, 227
  toroidal, 226
Radiation resistance, 238
Rat race, 136
Reactive loads, 50
Reactive stubs, 49
Reactivity ratio, 405
Reciprocity theorem, 488
Reflection coefficient, 39
Relative phase velocity, 33
Resonance condition, 163
Rotary couplers, 66
Rotary joints, 66
  coaxial, 66

S

Scanning, 459
  amplitude, 459
  electronic, 467
  frequency, 459
  phase, 459
Shorted stubs, 51
Sidelobe level, 235
Slotted line, 51
Slot types, 363
Slow-wave structure
Smith chart, 52
Space factor, 287
Standing wave ratio, 45
Symmetry,
  electrical, 110
  geometrical, 110
  physical, 110

T

Transmission-line charts, 52
Transmission-line matching,
  broadband, 55
  by lumped reactances, 56
  by quarter-wave transformer, 55
  narrowband, 54
Transmission lines, 18
  basic types of, 22
  coaxial, 24
  dielectric-rod, 33
  efficiency of, 45
  electrical length of, 48
  homogeneous, 18
  impedance transformation in, 48
  microstrip, 27
  microwave, 63
  open, 18
  optical-fibre, 34
  strip, 26

surface-wave, 32
uniform, 18
wire, 22
Travelling wave ratio, 44
Trombone, 182

U

U-bend balun, 292
Uda-Yagi antenna, 290, 362
Umov-Poynting theorem, 484

W

Waveguide bends, 74
Waveguide corners, 74
Waveguide diaphragms, 76
capacitive, 77
inductive, 76
resonant, 77
Waveguide posts, 78
capacitive, 78
Waveguide twists, 74
Waveguides, 18
breakdown in, 47
butt joints in, 75
circular, 29
double-ridged, 30
rectangular, 29
single-ridged, 30
Waves, 19
E, 19
electrical, 19
H, 19
hybrid electromagnetic, 19
incident, 36, 83
magnetic, 19
reflected, 36, 83
transverse electromagnetic, 19

Y

Y-circulators
stripline, 210
waveguide, 209

Z

Zones, 220
far, 220
near, 220
transition, 220
Zoning, 431

To the Reader

Mir Publishers welcome your comments on the content, translation, and design of the book.

We would also be pleased to receive any suggestions you care to make about our future publications.

Our address is:
USSR, 129820, Moscow, I-110, GSP, Pervy Rizhsky Pereulok, 2, Mir Publishers